DOCUMENT RESUME

ED 045 169 LI 002 388

AUTHOR	Mikhailov, A. I.; Giljarevskij, R. S.
TITLE	An Introductory Course on Informatics/Documentation.
SPONS AGENCY	International Federation for Documentation, The Hague (Netherlands).; United Nations Educational, Scientific, and Cultural Organization, Paris (France).
REPORT NO	COM-WS-147
PUB DATE	70
NOTE	202p.
EDRS PRICE	EDRS Price MF-$1.00 HC-$10.20
DESCRIPTORS	*Documentation, Foreign Countries, *Information Dissemination, Information Needs, Information Retrieval, *Information Systems, *Information Utilization, *Technical Reports
IDENTIFIERS	FID, *Information Science Education, Information Scientists, International Federation for Documentation, Russia, UNESCO, United Nations Educational Scientific Cultur Organ

ABSTRACT

In 1963, the International Federation for Documentation (FID) promoted the institution of documentation and scientific information chairs at the universities and other institutes of higher learning in all its member countries. A chair of scientific information was set up at the Lomonosov University of Moscow in the 1963/64 academic year. One of the purposes of the chair is to instruct students in the basic techniques and acquaint them with the main facilities used in scientific information work. A 24-hour optional course in "informatics" is offered to students in the natural sciences; this guide was written on the basics of this course. The guide contains materials for an introductory course in informatics/documentation science including: (1) lectures, (2) checkup questions, (3) tests, (4) lists of references, (5) curriculum and syllabus of the lectures and (6) practical lessons. The guide is intended for students of higher learning in countries which lack instruction in the discipline; it can also be used for the initial training of information officers in developing countries. (Author/NH)

ED0 45169

COM/WS/147

UNITED NATIONS EDUCATIONAL,
SCIENTIFIC AND CULTURAL ORGANIZATION

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

AN INTRODUCTORY COURSE
ON
INFORMATICS/DOCUMENTATION

by

A. I. Mikhailov and R. S. Giljarevskij

LI 002388

This draft text for an introductory course in documentation science is reproduced as a working document in the English version submitted by FID.

PREFACE

The scientific and technical revolution that is under way in the world of today is associated with a precipitous growth of scientific information. The growing amount of scientific and other publications is the most conspicuous feature of the phenomenon which is frequently described as "information explosion". Scientists and engineers naturally strive to get as much information as possible in a form convenient for use and spending minimal effort. In response to this social need, a special kind of activity has emerged, called scientific information, or documentation. Most countries have created ramified networks of information centres or documentation services which regularly provide the specialists with information they need.

It is a specific feature of information activities that an overwhelming majority of workers in science and technology take part in them. They record the results of their research, design and development work in papers, books, reports, projects, invention specifications, and these are later used by other scientists and engineers as initial data for the further research and development endeavours. There is a special category of scientific workers - information officers - whose task is to regulate the flows of these documents and to mould the totality of scientific facts and data into forms convenient for use. The efficiency of scientific information activities depends on the availability of highly skilled and experienced personnel, on the one hand, and on the acquaintance of all scientists and engineers with the essentials of informatics and their ability to find and use scientific information accumulated through the ages by mankind's exploring endeavour, on the other hand.

In many countries such training is provided to information officers and users through regular college and university courses. In 1963, the International Federation for Documentation (FID) addressed to all its member countries an appeal to promote the institution of documentation and scientific information chairs at the universities and other advanced learning establishments. In particular, a chair of scientific information has been set up at the Lomonosov University of Moscow in 1963/64 academic year. One of the purposes of the chair is to instruct the students in the basic techniques and to make them acquainted with the main facilities employed in scientific information work. A 24-hour optional course in informatics is offered to the students in the natural faculties; this Guide has been written on the basis of this course.

The Guide contains materials for an introductory course in informatics/documentation science: the text of the lectures, checkup questions for self-control, tests or examinations, lists of references suggested for further study, curriculum and syllabus of the lectures and practical lessons. The Guide is intended primarily for the students of higher learning establishments in the countries as yet lacking any regular instruction in the discipline; it can be also used for the initial training of information officers in the developing countries.

The Guide was written under a Unesco/FID contract. The authors wish to acknowledge their deep gratitude to the FID Secretariat and the staff of the Unesco Department of Documentation, Libraries and Archives for valuable advice and assistance rendered during the preparation of the manuscript. The principal notions and ideas expounded in the Guide have been formulated by the authors in the course of writing a monograph on the fundamentals of informatics ("Osnovy Informatiki", Moscow, Nauka Publishers, 1968, 766 p.) conjointly with A.I. Chernyi to whom they offer their sincere thanks for his materials for the university lectures that he had placed at their disposal and the helpful ideas suggested by him in discussing those lectures. Thanks are also due to E. Azgaldov and K. Erastov who translated this work into English.

The authors are hopeful that their Guide will be found useful not only by the students being initiated into scientific information activities, but by their teachers as well, and they would be grateful for any comments or suggestions.

CONTENTS

Page

Preface 1

1. Scientific Information Activities - Organic Part of Scientific Work 7

2. Scientific Literature - Source of Knowledge and Means of its Dissemination 19

3. Information and Bibliographic Publications - Sources of Data on Literature 41

4. Information Agencies and Special Libraries - Auxiliary Institutions of Science 63

5. Essentials of Information Retrieval 79

6. Conventional Information Retrieval Systems 89

7. Descriptor-Type Information Retrieval Systems 107

8. Technical Means of Information Retrieval 123

9. Reproduction of Information Documents 137

10. Utilization of Scientific Information 151

Syllabus and Programme for an Introductory Course on Informatics/Documentation 165

Figures 1 to 40 . . . . 169 - 208

LIST OF FIGURES

See page

1. Growth of world science periodicals 8
2. Tasks and stages of scientific-information activities 11
3. Types of scientific documents and publications 21
4. Citations of science documents in later publications 34
5. Distribution of new references to earlier publications in a hypothetical field 35
6. Scatter of citations in an isolated special literature 36
7. Half-life of scientific literature 37
8. Age distribution of papers cited in publications of 1965 37
9. Two different abstracts of the same paper as abstracted in the VINITI Abstract Journal "Chemistry" and the "Chemical Abstracts" 49
10. Arrangement of entries for the same books in a classified, alphabetical subject and author order 51
11. Basic types of bibliography 52
12. Layout of entry in a permuted-title index (KWIC Index) 57
13. Portion of a page in the Science Citation 59
14. Channels of dissemination of scientific information from the generators to the users 63
15. Distribution of scientific and technical information flows among the information agencies in the USSR 67
16. Diagram of information retrieval 80
17. Dependence of recall on precision 83
18. General block-diagram of an information retrieval system 84
19. Arrangement of bibliographical particulars on the title-page and in the entry 90
20. Scheme of rules for description of the basic types of publications . . 92
21. Specimen index entries 92

See page

22. Graph of an hierarchical classification and the relationship between its divisions 96

23. Main divisions in the Universal Decimal Classification (U.D.C.) . . 98

24. Main divisions (continued), and synopsis of auxiliaries in the U.D.C. 98

25. Deciphering of a U.D.C. number 98

26. A hypothetical classification of objects according to attributes which do not form natural hierarchy 108

27. Graphical representation of relations between sets 110

28. Working principle of the Uniterm System 114

29. Specimen layout of a subject class in a thesaurus 118

30. Examples of codes for edge-punched cards 127

31. The working principle of body-punched and superimposable cards . . . 128

32. Aperture cards 130

33. Microfiches (specimens) 130

34. Schematic operation and main units of Filmorex system 131

35. Schematics of the basic copying processes 140

36. Relative advantages and disadvantages of the different copying processes 143

37. Schematics of the different printing methods 144

38. Block-diagrams of the different printing machine types 144

39. Schematics of operation of the hectograph, rotary stencil duplicator and rotaprint 147

40. Choosing means of document duplication 149

1. SCIENTIFIC INFORMATION ACTIVITIES - ORGANIC PART OF SCIENTIFIC WORK

Science is a form of social consciousness that enables Man to learn the objective laws of Nature and Society and to put his knowledge into practice. Science, too, is governed by its own laws of development which it is necessary to know for successful work in any particular field of research.

Laws of the development of science

One of the principal laws is the differentiation of science.

Soviet Academician A.N. Nesmeyanov wrote in this connection: "Every field of science differentiates and gives rise to ever new disciplines. Chemistry, for instance, was divided already in the last century into organic and inorganic, analytical and, later, physical chemistry. In this century, physical organic chemistry has emerged, and quite recently the element-organic chemistry has become a discipline in its own right. In physics, the fields that became established as independent disciplines using research methods of their own were molecular physics, optics, acoustics, solid state physics, and - in the last decade - nuclear physics, etc.(1).

Differentiation and specialisation of science help the scientists penetrate ever deeper into the mysteries of Nature and Society, accumulate facts, and establish the interrelation of phenomena. This process, however, is fraught with a danger of which scientists have long been aware. Nearly a century ago the famous German chemist R. Wirchow wrote: "We all of us who describe ourselves as explorers of Nature actually have mastered only some particular field of it ... In all the other fields we are all half-ignoramuses."(2) In effect, differentiation leads to an ever greater insulation of scientists who are gradually losing orientation in the general system of scientific knowledge and even within their own discipline find it increasingly harder to keep *au courant* with their colleagues' work.

At the same time, it is the interaction and interpenetration of sciences that are particularly fruitful. Important trends of research are apt to arise at the junctures of related, and sometimes even of remote sciences. This regularity of the evolution of science has been pointed out by F. Engels who in his *Dialectics of nature* rightly predicted that it was precisely at the points of

---

1. A.N. Nesmeyanov, "Na styke nauk" (At the juncture of sciences), *Nauka i Zhizn*, 1958, no. 3, p. 1 /The VINITI system of transliteration is used throughout this volume/.

2. R. Wirchow, *Die Freiheit der Wissenschaft im modernen Staat*. Berlin, 1877, p. 13.

contact of sciences that the greatest results were to be expected.(1) The emergence and development of biochemistry, biogeochemistry, biophysics, physical chemistry, bionics, cybernetics, mathematical linguistics, engineering psychology and other disciplines are sufficient proof of this statement. It is evident that to ensure interaction of sciences in the environment of their continuing differentiation it is necessary to establish reliable channels of communication between the scientists working in different disciplines and trends of research.

Another important regularity in the development of science is its historical and international continuity. If every scientist and the scientists in every country and epoch had only themselves to accumulate the knowledge they need and had to rediscover laws, science would hardly make any headway. The accomplishments of scientists the world over and of the previous generations constitute the groundwork for the contemporary scientists to base their work on. K. Marx wrote: "Universal labour is all scientific labour, all discovery and all invention. This labour depends partly on the co-operation of the living, and partly on the utilization of the labours of those who have gone before".(2)

It is natural, therefore, that every generation of scientists is engaged not only in the origination of new scientific data, but also in special activities of classification, evaluation and generalization of the data earlier accumulated by their colleagues so as to make these data accessible not only to the contemporaries, but to the future generations of scientists as well. Obviously, this accessibility largely depends on the form in which scientific knowledge is disseminated among the contemporaries and handed down to the posterity.

Theory and methodology of science are beyond the scope of these lectures. These problems are the concern of a new scientific discipline, named science of science. We are not going to consider all the regularities of science, but shall only dwell on those of consequence for our subject. However, there is one more regularity that deserves to be pointed out here. We mean the accelerated growth of science.

It has become a newspaper and radio cliché that the world has entered an epoch of scientific revolution, a powerful explosion of scientific research. Every schoolboy nowadays knows that science is developing at a gigantic pace and that before he has time to go over to the next grade in his school, developments will take place in science which earlier had taken decades and centuries to accomplish.

It may be asked: how can one measure the growth of science? Aren't some major discoveries more valuable than hundreds of minor ones? Indeed, so far in making quantitative assessments of the rates of scientific development we can

---

1. F. Engels, Dialectics of nature. 3rd rev. ed. Moscow, Progress Publishers, 1964, p. 297.

2. K. Marx, Capital. A critique of political economy. C. Vo.3, book 3, Ed. by F. Engels. 3rd imp. Moscow. Progress Publishers, 1966, p. 104.

only lean upon such indirect indicators as allocations to scientific work, number of employees in the sphere of science, or the number of scientific publications. Howsoever approximate the data upon which such measurements are based, they are sure evidence that each of these indicators grows by equal number of times in equal time periods.

Let us illustrate this by examples. In 1800, there were about 1,000 scientific workers in the United States; by 1850, their number rose to 10,000; by 1900, to 100,000; and presently their number is in excess of 1,000,000. In 1800, there were about 100 journals in the world; in 1850, there were 1,000; in 1900, 10,000; and by the 1950's, close on 100,000. In both cases a 10-fold increase took place after every fifty years. In mathematical terms such growth is described by an exponential dependence and can be graphically shown by an exponent curve.

American historian of science Derek de Solla Price has drawn up a table showing that while the world population doubles every 50 years, the percentage of new scientific workers per 1,000 inhabitants doubles every 20 years, the number of qualified scientists and of scientific journals, every 15 years, and the literature devoted to certain subjects, for example, X-rays and experimental psychology, every 10 years. He made quite a clear picture of the rates of the growth of science in the following words: "To put it another way, using any reasonable definition of a scientist, we can say that 80 to 90 percent of all the scientists that have ever lived are alive now. Alternatively, any young scientist, starting now and looking back at the end of his career upon a normal life span, will find that 80 to 90 percent of all scientific work achieved by the end of the period will have taken place before his very eyes, and that only 10 to 20 percent will antedate his experience".(1)

The accelerated growth of science accompanied by frequent (every 15 to 20 years) doubling of scientific literature has been long since causing apprehension among scientists. Written records - journal articles, books, and other publications in which the scientists describe the findings of their research - are the basic means for the conveyance of scientific knowledge in time and space. Until quite recently, it has been the system of written records and publications and the simultaneously emerged system of library and bibliographic services provided for scientists, that had been ensuring the historic continuity, internationality and accelerated growth rates of science. At the present stage, however, this system no longer satisfies the scientists.

The apprehension was voiced in such well known utterances as that by the late USSR Academy of Sciences President S.I. Vavilov who visualized the picture of a scientist standing perplexed before the Himalayas of libraries unable to extract from them the grains of gold he needs, (2) or by the English physicist

---

1. D.J. de Solla Price, Little science, big science. New York - London, 1963, p. 1-2, 6-7.

2. S.I. Vavilov, Neskol'ko zamechanij c knigakh. (Some notes about books). Sovetskaja kniga, 1967, N 1 p. 15.

J.D. Bernal, who said that it is sometimes easier to rediscover a phenomenon than to learn from literature that it has been discovered earlier.(1) Scores of similar statements by less prominent scientists can be cited as well. Here is a word by the well-known French physicist Louis de Broglie: "A scientist often feels himself buried under the heaps of articles and monographs appearing all over the world; for all the assistance rendered by the bibliographies, he as often as not cannot manage to read them through, to say nothing of thinking them over. Flooded by the unceasing stream of publications, he is under the constant risk of getting mazed in trifles and missing the important things".(2)

Social demand for scientific information activities

Let us take a look at the obtaining situation from yet another standpoint. Since every scientist who wants to work efficiently must know what has been done in his field earlier and what the other scientists presently working on the same or related subject are doing, he has need spend part of his time searching for this information. Surveys show that many researchers in the natural sciences spend on this one-third or even half of their working time. This naturally means a reduction of the time spent on research proper and hence of efficiency of scientific work. And even so a scientist can never be sure that he has looked through all the relevant material.

The results are deplorable: quite a number of research projects are conducted in vain since they are mere repetitions of what has been done earlier by other scientists. Huge amounts of material and intellectual resources go down the drain.

All these facts were adduced with the unique purpose of showing that at the present stage of science development there has arisen a social demand for a special type of activity in order to provide the scientists and engineers with the knowledge, or, as we shall hereafter call it, scientific information they need. And such scientific information activities did evolve. There are thousands of large and small establishments in the world called information or documentation centres and services or institutes, bureaux, and departments of scientific and technical information. Ramified networks of such establishments providing regular assistence to research and development groups and teams are functioning in the majority of countries. The basic principles underlying the organization of such networks and the activities of the main information establishments in the Soviet Union will be described later on.

We must now give a clear notion of some general propositions pertaining to scientific information activities. First of all, it should be clearly understood that these activities have emerged as an altogether natural phenomenon in the course of the development of science. In effect, differentiation and specialization in science refer not only to the division of the fields and trends, but also

---

1. J.D. Bernal, Science in history. London, Watts, 1954.

2. L. de Broglie, Sur les sentiers de la science. Paris, Michel, 1960, p. 340.

of the types of work performed by the different scientists. It is not so very long since each scientist used to do everything required for his research himself. He conducted observations, proposed hypotheses, tested them in experiments, searched for the data required and classified and generalized these.

As time passed, experimental techniques became so involved and the methods used came to require the application of such sophisticated mathematical tools that the experimentation was established as a speciality in its own right. This process took place at different periods in different sciences, and it is not yet completed in some of them. As Soviet Academician P.L. Kapitsa worded it: "We know full well from the history of physics that the division of the physicists into theoreticians and experimentalists has taken place quite recently. In the past, not only Newton and Huyghens, but even such theoreticians as Maxwell used themselves to test their theoretical conclusions and constructs. Nowadays it is only in exceptional cases that a theoretician has to stage experiments to test his theories".(1)

In the last decades the turn has come for scientific information searching. Scientific information activities came to be established as an autonomous speciality in the course of the social division of scientific labour. A part of scientific workers now devote all their efforts to the gathering, critical analysis and generalization of all the available scientific knowledge pertaining to a certain subject field or a scientific discipline. Scientific information activities today require special knowledge on the part of scientists and application of complex techniques and equipment. The task has no longer to be shouldered by individual scientists, as there are specialized teams of scientists to cope with it.

Tasks and stages of scientific information activities

What are the tasks facing information activities and into what stages it is accordingly broken up?

1. First of all, it is necessary to gather all the required scientific information with the maximum exhaustivity possible. This means that from the collosal amounts of scientific knowledge one must select and regularly update all and only the information that might be of use for scientific or practical work in one or several disciplines, one trend of research, or even a single R & D team or establishment or a group of these.

2. The next step is analytico-synthetic processing of scientific information. This means that every group of data recorded in a scientific document is to be analysed from the standpoint of its contents while the document is to be analysed from the standpoint of its form. The results of such analyses are briefly written down or synthetized with the aid of special notation developed for the purpose of scientific information and based on the words of natural

---

1. P.L. Kapitsa, Teorija, eksperiment, praktika. (Theory, experiment, practice). Moscow, 1966, p. 16.

language, letters, figures, formulae, tables, and diagrams. Information thus processed becomes suitable for subsequent storage, retrieval and use in scientific work.

3. The following task is long-term storage of scientific information. The objectives here are to ensure that the data gathered take as little storage space as possible, and that the media upon which it is recorded provide for a longest term of storage and be arrangeable into an orderly system enabling fast and easy retrieval of all relevant data.

4. Information retrieval - which is a major aspect of information work - consists in a series of logical operations ensuring the finding of the information a scientist needs. Any procedure used to accomplish that task must be such as to allow at every point to retrieve from a collection of data all and only those providing a direct answer to the question posed by any member of the user community.

5. Since the results of each of the above processes and of information retrieval first and foremost may be of interest to a great number of scientists and practicians simultaneously, the task of scientific information dissemination crops up. Scientific documents (or parts thereof) carrying the relevant information must be reproduced to obtain the sufficient number of identical copies for the speedy dissemination among all these interested.

6. Finally, it is not unimportant to scientific information activities how their results are used. For that reason the objectives of these activities include the popularization of the accomplishments of science and technology and of the most efficient techniques of handling scientific and technical documents, the study of the efficiency of the utilization of these documents, and of the impact exerted by information work upon science, technology, and the national economy as a whole.

The fulfilment of the enumerated tasks and the performance of the corresponding stages of work require the elaboration of the theory, methodology, organization and the technical means of scientific information activities. These are the functions that constitute the subject of a new scientific discipline that we propose to call "Informatics".(1)

Basic concepts

Before we proceed with our exposition, we must introduce and define some basic notions. These are first of all the concepts of "information" and "scientific information" as well as the concepts derived from them: "scientific information activities", "informatics", "information officer" and "information scientist".

---

1. A.I. Mikhailov, A.I. Chernyi, R.S. Giljarevskij, "Informatics - new name for the theory of scientific information". FID news bulletin, 1967, 17, N 7, p. 70-74.

Information is certain knowledge, a totality of some data, and known facts.(1) In a broader, philosophical sense information can be defined as the content of the relation between interacting material objects which manifests itself in a change of state of these objects. Since all material objects can be divided into organic, inorganic and thinking objects (human), three kinds of information are accordingly distinguished: elementary, biological, and logical (semantic) information. Logical information is proper only to human society and its contents are ideas and images. Semantic information is actualized through man's oral and written speech.

Scientific information is logical information obtained in the process of cognition adequately reflecting the laws of the objective world and used in the socio-historical practice.

The term "scientific information" is, therefore, to be taken as generic: the attribute "scientific" by no means stands to indicate that scientific information had need be the result of purely scientific activity. The specific concepts will be by the different scientific disciplines - biological, physical, chemical, etc. information, and by the fields of practical activity - technical, agricultural, medical, political, managerial, etc. information.

Scientific information activities is a special part of scientific labour, which has become isolated in the course of the social division of this labour. Its purpose is to provide the scientists and practical workers with all the information they need by means of gathering, analytico - synthetic and logical processing, storage, retrieval and dissemination of scientific information.

A specialist in a field of science or social practice whose sole occupation is information work is usually called an information worker, or officer.

As you will have a chance to learn later on, some functions and processes of scientific information activities are akin to librarianship and bibliographic work, which have existed for thousands of years. In effect, libraries, too, collect literature process and store it and serve their patrons. This fact has given rise to the widespread opinion that scientific information activities have emerged to supplant the library art, since libraries allegedly are no longer able to serve science in our days. We shall devote a special lecture to a comparison of the activities of the information agencies and libraries. We should only like to warn rightaway that the above opinion is too straightforward and therefore not altogether true, be it for that reason alone.

Scientific information activities are to embrace primarily that part of scientific labour which had been formerly done by every scholar on his own. Of course, scientists and practicians have been making use of libraries and bibliographic aids in their research work since the days of yore. As scientific information activities are developing, this aid, contrary to losing its importance, is even to expand in scope and become more complex. That is why we are

---

1. Philosophical dictionary, Moscow, 1963, p. 172.

going to pay in our lectures due attention to those aspects of library work and bibliography that have retained their importance for meeting the information needs of the scientists.

Informatics is a scientific discipline which investigates into the structure and properties (and not the specific contents) of scientific information, as well as the regularities of scientific information work, its theory, history, methodology, and organization.

The purpose of informatics consists in developing optimal methods and means of presentation, recording, collection, analytico-synthetic processing, storage, retrieval and dissemination of scientific information. Informatics deals with semantic information, but is not involved in qualitative estimation of this information. Such an estimation can be carried out only by specialists in the particular fields of scientific or practical activity.

It should be mentioned that the term "informatics" is quite novel and has not yet received a common acknowledgement. It has been suggested by the authors of these lectures jointly with A.I. Chernyi to replace a number of existing terms that we deem unhappy. The oldest and most common of these is "documentation". It has been proposed for the designation of a specific activity of collection, processing, storage, retrieval and dissemination of documents as distinct from librarianship and bibliography back in 1905 by Paul Otlet, the well-known Belgian scientist and public figure. In 1934, it was used in the name of the Institut International de Documentation organized on the basis of the Institut International de Bibliographie which was established in 1895, and in 1937, in the name of its successor, the Fédération Internationale de Documentation.

In Soviet special literature a related term has been used rather frequently - "documentalistics".

We believe that both these terms - "documentation" and "documentalistics" - are inadequate, for they lay the stress on documents whereas the subject studied by the discipline in question is the structure and properties of scientific information. Nor can we call adequate the term "theory of scientific information", which is also sometimes used in that meaning, for the discipline embraces not only the theory, but also the methodology, history and organization; moreover, what is meant by "scientific information" in that term is in actual fact "scientific information activities". All these reasons prompt us to urge the introduction of a new term - informatics.

Specialists in this new discipline - informatics - likewise must have a name. Our suggestion is "information scientists".

Interrelationships between informatics and other disciplines

Informatics, being a newly emerged scientific discipline, is based on numerous other disciplines and makes use of their methods. We must say at least a few words on its related sciences that are quite diverse: mathematical information theory, cybernetics, semiotics, linguistics, psychology, library science, bibliography, book science, science of science, as well as technical disciplines.

The concepts which are central to the mathematical information theory are the concepts of information and of the measure of its quantity, while its major objective is to investigate into the process of the transmission of information through a communication channel. This theory derives its definition of information from statistical reasoning and completely disregards the contents, or semantic value, of the messages transmitted. Nevertheless, the process of storage, retrieval and dissemination of information, considered in sufficiently general terms, may admit of being treated as information transmission through a communication channel. This explains why informatics makes so wide a use of a great many notions of information theory.

Cybernetics has been defined as the science "of the modes of perception, storage, processing and use of information in machines, living organisms, and associations of these".(1) Characteristically, there is an almost verbal coincidence of the opening part of this definition with the stages of scientific information activities described above. Another factor of the rapprochement between informatics and cybernetics is the circumstance that at present the efficient solutions of the main problems of informatics - retrieval and logical processing of scientific information - are based on electronic digital computers, or cybernetic devices.

Semiotics, or the general theory of symbolic systems, which is sometimes regarded as a part of cybernetics, is currently being developed at the interface of mathematical logic and structural linguistics. Even the little thus far told on the analytical and synthetic information processing will suffice to make clear the importance of artificial formalized languages with special symbolisms or of the programming-type languages for informatics. The wide scope of subjects studied by semiotics, includes such aspects important for informatics as the principles of construction of artificial languages, the procedures of translation from a natural language into an artificial one and backwards, and the notation systems used to record scientific information in the different disciplines. Methods of mathematical logic are likewise employed by informatics for formalizing the logical inference procedures applied in the different sciences; this would provide for the automation of a great number of functions of scientific information activities.

Linguistics, which is concerned with the study of the regularities of evolution of the natural languages, shares much common ground with semiotics as regards the machine translation problem - a problem of great importance for informatics. Other chapters of linguistics, such as the general theory of translation, the principles of transcription and transliteration, and the principles of the construction of terms, too, find definite applications in informatics.

Psychology, particularly one of its recent trends - engineering psychology, places at the disposal of informatics very valuable methods of research to investigate into the optimal forms of scientific information tailored to the

---

1. A.N. Kolmogorov, "Kibernetika" (Cybernetics). In: Bol'shaja Sovetskaja Entsiklopedija, 2nd ed., v. 51, Moscow, 1958, p. 149.

convenience of perception, the utilization of scientific information, the efficient reading techniques, the construction of man-machine systems and a great deal of other problems associated with the human aspects of information activities.

Library science, which studies the essence, organization and methods of public use of written and printed records has historical ties with informatics. For thousands upon thousands of years the techniques of information activity individually carried on by every scientist as part and parcel of his work had been his personal affair. They were never generalized and at best passed down directly from teacher to pupil. The only social institution where these techniques could find at least an indirect reflection was the library. Library science has managed over the last century and a half to generalize some of these techniques. Although as yet it remains a descriptive science and certainly lags behind in solving the pressing problems of the servicing of science, informatics does make use of some of its results.

Bibliography is "the field of knowledge and of practical activity which has for its object the registration, description, classification and qualitative analysis of publications and the compilation of various aids which help to take one's bearings in the current literature, popularize it and promote its efficient utilization."(1) The scope of bibliography is much broader than that of library science and encompasses virtually all the aspects of human activities associated with written and printed records. Alongside with scientific and industrial purposes, it provides for educational and training objectives as well. However, supplying information on printed and written records to scientists and all kinds of practitioners remains one of the most important tasks of bibliography, and in this respect it is akin to informatics and information work.

Book science is a comprehensive discipline that studies printed and written records on a theoretical and historical planes. Since scientific information is at present disseminated in the form of publications and manuscripts, a great many aspects of book science are of significance for informatics, specifically the history of books, book statistics, publishing art (in particular, the theoretical and practical aspects of editing), and the graphic arts industry.

Science of science (organization of scientific labour) is an altogether recent scientific discipline which investigates into the regularities of the development of science and elaborates on this basis the principles of organization, planning, financing and management of science. Some of the problems explored by the science of science are important for informatics as well. These include the problems of the optimal system of scientific publications, the improvement of the efficiency of scientific labour, prognostication of the development of science, assessment of its level and development rates, and a number of others.

---

1. E.I. Shamurin, Slovar' knigovedcheskikh terminov. (Dictionary of library science). Moscow, 1961, p. 29.

Finally, the technical science provide for the elaboration of new technical means which is a sine qua non for enhancing the efficiency of information activities. It is difficult even to enumerate here all the fields of engineering that have to be made use of in developing special information devices. Microphotographs and records on various magnetic carriers are used for long-term storage of scientific information. Photography and electrical engineering, optics and electronics participate in the efforts to solve this problem. Considerable successes have been scored in scientific document reproduction techniques. Some new processes, for instance, electrography, considerably enhance the efficiency of information dissemination. As concerns information storage and retrieval, great hopes are set on the electronic digital computers. Engineers are presently striving to build specialized input and output devices for computers so as to provide for using them in automating the main routines of scientific information work.

So much for the range of sciences akin to informatics in its main pursuits. This range is of course even broader, but this general outline is enough to indicate the numerous relations of informatics with the present-day science and to delineate its scope of research.

On the contents of these lectures

We want to conclude with a word on our lectures. They should not be taken as a regular course expounding the fundamentals of informatics. Their purpose is only to serve as an introduction to this discipline.

Both the structure and contents of this course are determined by that particular purpose. We shall first of all make you acquainted with the existing system of scientific publications, primarily those providing aid in searching for the pertinent scientific literature. We shall speak of the activity of those establishments - scientific information agencies and special libraries - that function as middlemen between the scientist and his literature and as treasuries of scientific information.

The central place in the lectures is given to the essentials of information retrieval. Both conventional and unconventional systems as well as the facilities for their physical implementation are considered. Major attention is given to the simplest means which can be available to any student.

The course ends with a lecture on the means and methods of the utilization of scientific and technical information and the personal information work of the individual scientist.

Every lecture closes with checkup questions and references for further study.

Checkup Questions

1. What are the main regularities of the development of science?
2. What are the indicators that characterize the accelerated growth rates of scientific progress?
3. Which social demands have given rise to scientific information activities?
4. What is scientific information work, what stages does it consist of, what are the problems it strives to solve?
5. What is scientific information?
6. What is informatics?
7. What scientific disciplines is informatics related to?

Literature

1. A.I. Mikhailov, A.I. Chernyi, R.S. Giljarevskij, "Informatics: its scope and methods". In: On theoretical problems of informatics. FID 435. Moscow, All-Union Inst. for Scientific and Technical Information, 1969, p. 7-24.
2. H. Borko, "Information science: what is it?" American documentation, 1968,
3. Ch. P. Bourne, Methods of information handling. New York a.o., Wiley, 1963. XIV, 241 p., ill.
4. D.J. de Solla Price. Little science, big science. New York-London, Columbia Univ. Press, 1963, 119 p.

2. SCIENTIFIC LITERATURE - SOURCE OF KNOWLEDGE AND MEANS OF ITS DISSEMINATION

"If the history of science and technology is considered from the standpoint of the level of means of information transmission and storage, three major epochs can be more or less clearly distinguished thus far: the first, and the longest one, when oral speech was the chief if not the only means of information transmission; the second epoch which began with the emergence of a class society ... is characterized ... by the addition of writing to oral speech; finally, in the third epoch - from the establishment of the bourgeois society up till now - the main means of storage and transmission of scientific and technical information was book printing".(1)

Stages of development of written records

Since the methods of storage and transmission of semantic information have been playing a major part in the development of human society, it will be interesting to trace back the evolution of these methods. Oral speech proved inadequate for the task of information transmission at a certain stage of development when the need arose to convey messages over large distances and to preserve them for prolonged periods of time. Written language made its appearance being a system of physical fixation of speech sounds and words using conventional signs with the aim of their transmission in time and space.

Every civilization, as it perfects the oral and written language in the course of its development, works out certain types of documents which, along with oral communication, serve to convey the knowledge about the surrounding world and the skills in this or that craft or industry. The oldest scientific documents that have been preserved till our day are the tablets inscribed with cuneiform writing from Mesopotamia which date back to the 4th millemium B.C. Treatises in astronomy, geography, history, commerce and law have been preserved since the 4th millenium B.C. (the Sumer culture and the Assyrian-Babylonian culture which flourished in the 2nd millenium B.C.). Along with small clay tablets 2.5 cm high, there were produced big plates 40 cm high which contained up to 400 lines written in 4 columns on either side. These books of clay served human cultural and scientific needs for thousands of years, but their major shortcomings were unwieldiness and insufficient capacity. Some texts took up to 10 tablets to record, and the collections numbering tens of thousands of tablets - remnants of such libraries have been preserved till our time - took enormous storage space to accommodate.

Papyrus scrolls were a more compact form of document which made it possible to accumulate vast collections of manuscripts. Beginning with the 3rd millenium A.D., this was the usual form of books in Egypt. In a scroll, the text

---

1. O.S. Kulagina, L.A. Kaluzhnin and V.V. Ivanov, "O kul'turnom znachenii mashinnogo perevoda". (On the cultural significance of machine translation). Vestnik istorii mirovoj kultury, 1961, N 3(27), p.23.

was arranged crosswise to the length in columns of 25 to 25 lines. The fragility and impermanence of papyrus are to blame for the scarcity of the specimens of ancient Egyptian documents preserved; the oldest of these dates from the 18th Century B.C. Mostly, these are the ritual "books of the dead" that the archaeologists unearth from the pyramids and other burial places. In Greece and Rome, scientific knowledge had been for a long period disseminated through oral speech. It is known that manuscripts came to be used for the purpose as late as after Aristoteles. Not a single original text has been preserved since the classic times, and it is very often that centuries separate an earliest preserved copy from the conjectured date of writing of the original text. What is known about Greek and Roman papyri is based on Asia Minor and late Egyptian specimens. The mean length of a scroll was at most 10 m, and the width, 30 cm.

Parchment made of calfskin has been known as material for writing since the 3rd Century B.C. It admitted of writing on both sides, and was preserved longer than papyrus.

It determined the transition to the codex, a modern book form. Papyrus codices were manufactured till as late as the first centuries A.D., but by the 5th Century they were ousted by parchment codices. The codex is a book form of a greater capacity than a scroll, it is better suited to the recording of long texts and to making references. Up till the 12th century, parchment scrolls were the only book form that existed in Europe. Their biggest drawback was a high cost, which prevented a broad spread of written documents. A whole herd had to be skinned to make a single copy of a parchment book.

Parchment was superseded by paper, which had been invented in China in 105 A.D. In the Middle East, its manufacture was started in 751 A.D., and in the 12th Century it penetrated into Europe through the Arabic conquerors of Spain. Paper books which were considerably cheaper than parchment books, came to be widely used for scientific and educational purposes. The several thousands years of the evolution of written documents show that the form of the latter has been changing under the pressure of the needs of the community: documents were becoming an ever more capacitous, convenient, and cheaper medium for information recording. This resulted in the emergence of the book form, or codex, which remains the main form of document till our day.

An ubiquitous need began to be felt in the 15th Century for a new technique of document production. Handwritten books were no longer sufficient for the cultural and scientific demands of the society. There were two basic reasons for this. Firstly, they were too costly and laborious to make. Secondly, the manual copying of texts could not provide for producing a sufficiently great number of identical copies, since the copiers frequently disfigured text. It was at this time that the German artisan J. Gutenberg invented printing with movable metal types. The technical means upon which book printing was based had been known since the antiquity - engraving and moulding techniques, and the wine press that Gutenberg modified to build his printing press. But it was only the urgent demand of the community for rapid and exact mechanical reproduction of texts that called into existence this invention which even now remains for us an object of unceasing admiration and the prototype of the techniques that we continue to use to produce printed documents.

Scientific documents and their types

Documents, in a broad sense, include not only inscriptions, manuscripts, and printed publications, but also works of art, numismatic items, and museum exhibits of mineral, botanic, zoological and anthropological nature. Any material object that is a record or corroboration of some knowledge and can be included in a certain collection may be regarded as a document. A scientific document is a material object carrying scientific information whereof it is a record designed for the transmission of this information in time and space, and utilized in the social practice.

Different types of scientific documents appeared in different epochs, and during the last centuries and even decades they have been undergoing substantial changes. Books exist for several thousand years already, patent specifications have been known for some five hundred years, scientific periodicals are but 300 years old, while journal articles in their present form are even younger - 100 to 150 years. Book science as yet lacks a well articulated typology of scientific documents. The distinction between published and unpublished documents until quite recently has been considered to be the basic division. Even a few decades ago ideas and facts were recognized to have entered the scientific usage only after their publication, which then meant a more or less wide spreading and official registration of the documents which described them. This discrimination is now of lesser consequence for scientific information activities, since, first of all, unpublished documents contain quite a lot of valuable scientific information which is ahead of that appearing in the publications, and secondly, the present-day reproduction facilities make the division into published and unpublished rather conventional. Such documents as scientific and technical reports, dissertations, translations of scientific materials, which are conventionally considered unpublished, are now quite often disseminated in hundreds and thousands of copies.

Informatics has moved to the foreground the division of documents into primary and secondary. This division, too, is quite conventional and approximative, as it largely refers not to the documents but to the scientific information they contain. It is accepted that the primary documents record the immediate results of scientific research and design and development activities, while the secondary documents deal with the results of the analytico-synthetic processing of scientific information contained in the primary documents. In actual fact, however, the historically evolved system of scientific documents is such that many of these contain both the results of genuine research and of the processing of information taken from earlier publications. Examples are many - articles in journals, monographs, textbooks, and last but not least, reference literature.

This division is convenient, nonetheless, as it allows to characterize the different flows of scientific documents in information work. We shall cling to it further on, taking the primary documents to be those documents and publications which mainly contain new scientific knowledge or new understanding of known facts or ideas, and the secondary documents to be those documents and publications which predominantly serve to give information on the primary documents. Taking into account the above reservations, one may include with primary documents and publications a majority of books (excluding reference literature, periodical publications, special types of technical publications, scientific and technical reports, dissertations, translations, information cards), and with secondary documents, reference literature, surveys, abstract journals, library catalogues, bibliographic indexes and card files.

Let us give a brief characterization of the main types of primary documents and publications. (We shall deal with secondary documents and publications in the next lecture).

Book is an ambiguous term. As concerns contents, one can speak of a book as a scientific, practical or literary (artistic) piece; if only the form is taken into consideration, then any printed or written reproduction of text or graphic figures may be taken to be a book. Informatics defines book as a nonperiodical many-page printed item. "The Recommendations on the International Statistical Indicators Characterizing the Publishing of Books and Periodicals", approved by Unesco in 1964, provide a clearcut distinction between books and pamphlets. Book is a nonperiodical printed publication of no less than 49 pages, excluding the cover and the title page. Pamphlet is a nonperiodical printed publication of a size no less than 5 pages and no more than 48 pages.

As mentioned, the major advantage of book printing over handwritten books is the capability of manufacturing any required number of identical copies exactly matching the original one. In addition, in most countries a printed book is subject to special registration at the moment of issuance, which ensures keeping an accurate record of all publications. These advantages have made it that books became and still are the most important type of scientific document playing a major role in the development of science and technology.

It was only in the recent years that scientists began to voice their dissatisfaction with the existing system of dissemination of scientific knowledge in which books occupy the key position. This discontent stems from a number of circumstances, including a long time lag between the writing of an original and its publishing, the inaccessibility of a great number of publications to the mass of scientists and experts, the difficulty of orientation in the ever growing book flow, etc. On the other hand, the principal features of printed records - black text against light-coloured background, the sizes of types,lines and pages, and the codex form continue to satisfy the most exigent reader today precisely like they did centuries ago, and perfectly conform to the faculty of the contemporary reader for mastering knowledge through reading. This grants the assumption that publications in their traditional forms will survive for quite a long time to come. As to the scientists' discontent with books, which is certainly justified in part - it means that the system of publication and dissemination of books should be improved.

The many operations involved in the manufacturing of a printed book take up a lot of time. Recalling that the publishing house and the printing office usually work on a great number of publications simultaneously, each of which is to be printed in thousands of copies, it will be clear why these would remain for months in press. This is one of the most serious drawbacks of books as a means of scientific information dissemination. Information contained in books often dates from years back and is of faded interest for the current needs of scientific research.

For that reason books weigh less in information work as compared with other scientific documents. And nevertheless a book of science is an important means of generalization of scientific knowledge. Books are the form of publication of major generalizing theoretical studies, and of explorations of crucial problems in science, economy and culture. They are an unrivaled means of education, teaching and upbringing of the new generations of scientists and practicians. The following book forms may be considered to be of prime importance for scientific information activities: monographs, collections of papers, scientific conference proceedings, textbooks and manuals, and official publications.

A monograph is a work of science or of earnest popular science that gives an exhaustive and comprehensive consideration to some problem,question, or subject. A monograph is usually written in accordance with a well elaborated schedule embracing all the aspects of the subject or phenomenon under consideration. It may be by one or several authors. The monograph was the predominant type of science book in the last century; it came successor to treatise which had prevailed in the literature of science in the previous epochs.

Presently monographs have ceded the priority to collections of papers, i.e. books consisting either of separate papers by the same or different authors, or of official, historical, archival or other documents and records. Collections of scientific papers written by different authors have acquired particular importance in the postwar years. In contrast to a monograph, a collection of papers, even compiled according to a definite schedule, needs not exhaust the subject, but may only touch upon some of its aspects. The collections which take less time to compile than monographs to write have come to be a most widespread type of scientific book, and the importance of the data they carry is so great that no specialist can dispense with them in his work.

Proceedings of conferences, congresses, and symposia are as a regular rule published after the closing of these meetings of scientists. Along with the resolutions adopted at the conferences, they contain the abstracts and sometimes complete texts of the reports delivered, as well as the summaries of discussions. More than a half of communications published in such books do not appear in other publications, e.g. journals. Hence their great value for scientific information work.

Textbooks and manuals hold a special place among other books used in scientific work. They describe but the principal facts and propositions and are mostly intended for instruction. However, the most original of them are of interest to the specialists, too, for the novel groupings and generalizations of known facts. Besides, textbooks of allied disciplines can be helpful to a specialist to familiarize with the range of problems in a field not of his immediate interest; they also often serve as standards of term usage.

Official publications are issued on behalf of organizations and institutions, and they are documents carrying information that bears a direct relation to the activity of these institutions which are responsible for this information. Many of the official publications have no scientific value, but some of their types will contain important scientific information, e.g. reports, plans and accounts describing the work of research and other institutions, their resolutions, decisions, and regulations, intra-agency norms (some of these are classed as special types of technical publications), industrial and engineering instructions, reference- information, methodological and statistical documents, training curricula and syllabi, etc.

As mentioned, all the enumerated types of books need not necessarily be in single volumes, but they can also be multivolume or be part of a series. A multivolume publication is a set of printed items contained in two or more volumes, books, parts or issues that make a distinct whole. A series consist of a number of independent published works usually characterized by common or close subject scope and issued by the same publisher under a common serial title. In contrast to the multivolume sets, the issues of a series may be unnumbered.

Over the past few decades major changes have been unfolding in book publishing that have affected also the sphere of science books. The so-called paperbacks - cheap pocket-size books printed in millions of copies - have been adopted for books of science, popular science and textbooks. This prompted some bibliologists to speak of a revolution in the books industry. The picture would be incomplete without a word on the number of science books. Unfortunately, we lack exact data on the number of books that have been published since the invention of book printing. By all likeness it can be roughly estimated at 30 million titles. There is more precision as regards the present annual book output of the world. According to Unesco, in 1952 it totalled 250,000 titles; in 1962, 385,000; in 1963, 400,000; in 1964, 415,000; and in 1967, 450,000. Considering that the books printed for sale constitute but a part of the world book output, one may suppose that the true figures would be the double of those cited above. Anyhow, the figures are impressive as they are. An annual 450,000 titles means a daily batch of 1,200 titles, or some 50 titles every hour, - in other words, a new title appears every minute round the clock.

According to some national statistics, science books make 1/5 to 1/4 of the total book market. One can thus speak of an annual increment of close on 100,000 new science book titles at least a half of which are new editions, that is not translations or reissues. We should, however, be cautious to accept wholesale all the complaints about the overabundance of books recalling that it is a natural source of anxiety to the scientists unaware of the lawfulness of the exponential growth of science.

A periodical publication is acceptedly a published work regularly appearing after determined intervals of time in separate issues of differing contents but under the same title. A periodical is usually meant to be published for an indefinite time. Usually its issues have identical makeup. Conventional forms of periodicals are journals and newspapers. Periodicals in a broader sense are sometimes supposed to include also the continued publications (serials), which are an intermediate form between books and journals.

The continued publications (serials) are collections of scientific papers and other documents put out by the various institutions, societies and other organizations, without any strict periodicity but in numbered issues and under the same common title /"Transactions", "Records", "Proceedings", etc./ and in identical makeup. It is sometimes hard to distinguish a continued publication from a book series or a multivolume set, on the one hand, and a journal on the other. The distinctive feature of the continued publications (serials) which helps to tell them from book series is an indication that, in general, they are produced by establishment other than commercial publishing houses. Serials are mostly published by academies, universities, scientific research and learning establishments, learned societies, etc.

Serials are distinguished from multivolume books in that they have no prefixed schedule and number of issues (volumes). Apart from publications issued without strict periodicity, publications which are issued at regular intervals but for a period longer than a year are also classed as serials (yearbooks and publications issued every two, three, four, five or more years). This distinguishes them from journals which are either published during the year at established intervals of time, or have a fixed number of issues per year.

Serials are an important source of scientific information: they often carry information which does not appear in the regular journals and print reports, papers and other articles in narrow subject fields related to the trends of research pursued by the given scientific establishment.

A journal or a magazine is a periodical publication regularly appearing in weekly, monthly, bimonthly, quarterly or semiannual issues, in identical makeup and carrying articles or other materials of scientific, technical or sociopolitical content, as well as works of fiction. Journals contain the current information and describe the recent advances in science and technology.

Scientific journals came into being 300 years ago. The precise date is recognized to be January 5, 1665, when the first issue of the French weekly "Journal des sçavants" appeared -- which was to give the name to this type of periodical publication. The main purpose of the Journal, which has predetermined the nature of scientific journals for a century and a half ahead, was to report on all the disciplines of science, literature and arts, with a special emphasis on the natural sciences and engineering. Direct quotations and excerpts were a regular method for rendering the contents of books at the time.

At first, science chronicles played a secondary part in the journal. Gradually, communications on the newly discovered natural phenomena and experiments in the natural sciences began to occupy more and more space. Original articles were rarely published in the journals in the 18th Century. For the most part these were by prominent scientists and had the traditional form of letters written by a scientist to his colleague that had been the conventional form used to communicate scientific discoveries in the preceding epochs.

Articles in scientific journals are presently the main source of scientific information and they hold a sure priority among all other types of scientific documents. A survey of the bibliographic requests submitted by several thousand American scientists has shown that 68 per cent of the requests were for journal papers. Special journals account for 70 per cent of all sources used by British electrical engineers.

Journal papers in natural, exact and technical sciences can be divided into three categories. The first category is original research papers which enable a scientist to reproduce the experiments and observations of the author of a paper and to verify his conclusions without any reference to other sources. The second category are preliminary publications and notes which possess certain features of scientific novelty, but are inferior to the first category of documents in respect of exhaustivity and accuracy of description. Lastly, the third category are the so-called explanatory articles that hardly carry any new scientific information. Their purpose is to analyse and discuss information published earlier.

On the whole, journal articles are the best elaborated existing system for communicating scientific information. This system, as we have seen, has taken shape relatively recently when it took over some of the functions previously fulfilled by science books. It is still developing and it requires a profound study. Recent observations show that gradually more and more articles in science journals are due to team effort. While 80 per cent of the articles at the turn of the century were by a single author, presently such articles make but less than a half. The number of articles by two, three, four and more authors has been growing steadily.

As compared with books, journal papers are published much faster. But they contain less generalizations and therefore soon become outdated. The obsolescence rate of both journal papers and books depends on the development rate of the fields of knowledge concerned.

The growth rate of the number of journal titles has been characterized in the preceding lecture - it is described by a tenfold increase every fifty years. While in the middle of this century there were 100,000 periodical titles, their number is to mount up to 1 million by its end. It is rather difficult to isolate from this total the number of science journals. For this we shall make use of a bibliographic reference aid entitled "Word List of Scientific Periodicals". Its first edition, published in 1924, recorded 24,000 titles. The third edition (1952), contained already 50,000, and the fourth edition (1965), over 60,000 titles of journals.(1) This latest edition records all science journals that came out between 1960 and 1965, including those no longer published. It has been roughly estimated that since about half of the number have ceased publication, close on 30,000 science journal titles are regularly being issued at present, with an annual total of 3 to 4 million articles. As we shall see, the latter are distributed very unevenly.

---

1. World List of Scientific Periodicals, Published in the Years 1900-1960, 4th ed. Ed. by P. Brown and G.B. Stratton. Vol. 1-3, London, 1964-1965.

The growing number of papers published in journals, the loosely defined profiles of these, the rapid obsolescence of the materials they carry, and the inadequate announcing of their issuing to scientists -- all this has contributed to the fact that since back in the 1930s the journal as a source of information has been target to criticism. In the past three decades, various projects have been advanced concerning the eventual replacement of journals by other media of knowledge dissemination. These projects are based on the idea to use for the purpose separate papers not published in periodicals but deposited in special branch-oriented centres and signalised by abstracting publications. One of the earliest projects of that kind was drafted by Soviet delegates to the 1933 International Geological Congress, and a similar proposal was made the same year by an American - Watson Davis. It was his proposal that J.D. Bernal used as the basis of his widely known plan put forward in 1939 in the book entitled "The Social Function of Science". The Bernal Plan was discussed at the Royal Society Scientific Information Conference in 1948.

Elaborating on his plan, Bernal suggested at the 1958 Conference on Scientific Information in Washington that the very form of science journal article should be renounced as the means for communicating R & D results. He argued that the form of a scientific paper does not provide for a rapid and adequate description of the entire course of research. The Bernal Plan is unlikely ever to be implemented since it did not take into account some recently discovered patterns in the science publications system. Apart from knowledge dissemination, journals have other functions to discharge, including upholding a critical atmosphere in science, priority protection, and assessment of scientists' competence. Still, the plan did contain useful ideas that are already being implemented in depositing of unpublished sci-tech documents.

Science books and journals remain the most important and time-honoured types of scientific publications and documents, and that is the reason why we have dwelt on them at such length.

Newspapers are defined as periodical publications that are generally issued daily, or one, two, or three times a week and carry information on current developments, primarily of socio-political nature. Newspapers also regularly publish popular science articles and sci-tech news features. The freshness and novelty of such newspaper pieces invest them with definite informative value. Newspaper clipping services provide for current awareness on items of topical interest as treated by the majority of a country's newspapers. However, the advertising tone so typical of many papers and the often unreliable information they give detract from their usefulness as sources of scientific information.

Special types of technical publications include standards, branch and interbranch technical documents, patent literature, technical catalogues and price lists for materials and equipment, as well as publications put out by information agencies. The term "special types of technical publications" is sometimes applied to unpublished scientific documents, too, such as scientific and technical reports, information cards, and translation of papers. The term has been coined and defined by Soviet librarian A.N. Barabanov who in mid-1930s identified the common features uniting these diverse publications in respect of contents, makeup, origins, and purpose.

Standards are scientifically substantiated technical documents which specify the types, sorts and grades of products as well as their quality, testing methods, packing, marking, transportation and storage. Standards also define the general technical values, measures, terms, and symbols. Usually they give comprehensive characterisation of the product concerned, but sometimes they deal with particular aspects, for instance the testing procedures.

Standards serve to ensure high quality of industrial and agricultural products and the unification of parts and assemblies as well as finished items. They promote the continually better utilization of materials, fuel and power, faster designing and lower costs of industrial products, and can serve as a source of information on technical accomplishments.

Cognate to standards are interbranch and branch technical documents (specifications, norms, and other instructional technical documents). These have a more restricted scope of application and are established either in the absence of a corresponding standard or in case of some special requirements for the product concerned associated with the peculiar manufacturing or operation conditions. Specifications are used on a wide scale side by side with standards. These are commercial documents that state the consumers' or buyers' technical requirements for products, materials, processes or services which the manufacturer is liable to ensure.

Patent literature is an important source of scientific and technical information. It makes it possible to trace back the history of any invention or discovery and to get a clear picture of the present state and development trends of technical thinking in the respective field of science or technology. Patents provide an opportunity for comparing and assessing the technological levels in the different countries and to ascertain the feasibility or otherwise of mechanisation and automation of various production processes. Timely acquaintance with patents prevents duplication of inventions or the development of techniques and machinery that already exist.

Author's certificates are the main form of authorship protection in the socialist countries. It warrants the perpetual preservation of an inventor's authorship, as well as his right to remuneration and all the privileges granted to inventors and innovators in those countries. In accordance with the structure of the socialist system under which all means of production are in the possession of the State, an author's certificate reserves for the State the exclusive rights of using the invention. In the capitalist countries, the inventors are granted patents whose holders are entitled to the utilization of the invention for 15 to 17 years.

The bulk of patent literature is made up of the specifications appended to patents and author's certificates. These consist of a brief description of the essence of an invention, indicate its scope of application, characterize the existing solutions of the problem concerned, and enumerate the drawings of the individual parts of the invention. Thereupon follows a complete description of the invention specifying the peculiar features and the essence of the new device, machine, product or chemical substance, and enumerating its main parts and assemblies with an explanation of their functions and interactions and references to the corresponding markings in the drawings. A patent specification ends up with the so-called "Claims", that is a brief summary of the object of invention underscoring the essential features which are regarded as the invention proper and are subject to legal protection.

Patent literature is very ample. The first patent was granted in Britain in 1449 and since the middle of the 16th century patenting of inventions has been a regular practice. The overall number of patent specifications published since is estimated at 7.5 million items, to which some 300,000 new titles are added every year. Periodical publications of the patent offices are likewise treated as patent literature. These are bulletins signalizing inventions, classifiers serving as aids in classifying inventions, and other publications giving information on patents and author's certificates.

Technical catalogues (also called industrial, commercial and company registers) are lists of products manufactured by industrial enterprises and concerns or sold by trading companies. These usually contain a specification of each product with drawings and pictures and sometimes also indicate the price, scope of application, and type or model. Technical catalogues are of three main types: the first type are those describing products of a particular kind or designation and manufactured by the different enterprises, trusts, and companies; the second type, by contrast, are those enumerating the different items produced by some one enterprise, trust or company; and the third type, reference catalogue, give comparative data on various kinds of similar products. The number of technical catalogues is so great that it is hard to keep account of them: their annual output is over half a million new titles. They are widely used by engineers, economists and scientists to study the up-to-date technical equipment with a view to its acquisition, economical utilization or improvement.

Information leaflets or newsletters are one example of the numerous publications issued by information agencies. They are circulated in series of several hundred copies, serving for information exchange between the research institutes, design bureaux of enterprises, and the design and development organizations, to prevent duplication and parallelism in their work. The leaflets, of 3 to 4 pages, briefly describe the engineering calculation techniques, novel procedures used in production, regulation, adjustment, testing and control of instruments and apparatus. Information leaflets deal with the particular technical solutions of a general significance and are published before the completion of an overall research or design project. They often contain the electrical or mechanical diagrams of the particular elements, assemblies and blocks, describe their design and cite the characteristics and properties of new materials.

Unpublished scientific documents, as we have seen, are a conventional category that includes very different documents: scientific and technical reports, dissertations, information cards, and translations. They all may be issued in a great number of copies and circulated as fullfledged publications. Another group of unpublished documents - manuscripts, proof sheets and preprints - are products of intermediate stages of the printing process. Let us dwell on some kinds of unpublished documents.

Scientific and technical reports (that is reports on the results of research and development projects) are a major source of information which is necessary for the development of science and technology, introduction of their accomplishments into the industrial practice, and improvement of engineering procedures. The techniques of writing reports are different in the different branches of science and technology, countries and even organizations. Nonetheless, certain common features are distinguishable in all these documents.

A scientific or technical report usually opens with an abstract or summary, briefly stating the task of the research and the results obtained. This is followed by an introduction giving background facts on the state of the elaboration of the problem under consideration both at home and abroad. The main text of the report contains the statement of the problem, the formulation of the technical assignment, an analysis of the existing methods of solution, a substantiation of the approach adopted, the necessary calculations and experimental results. The report closes up with conclusions that compare and analyse the theoretical and experimental data obtained during the research. The last section of the report contains an assessment of the results and their comparison with the technical assignment, as well as an outline of their prospective utilization. Usually a scientific or technical report is furnished with a review of the sources used. At times such a review is given in the first section of the report, in which case the latter is concluded by a subject index to the references.

This kind of document has become particularly widespread in the postwar years so that it is now vying with journals. The number of reports is also very great. In the USSR, some 5,000 research and development and over 8,000 design and development organizations carry out and describe in their reports up to 150,000 projects annually. In the USA, federal agencies alone issue over 100,000 reports every year. Since scientific and technical reports are not considered as publications and their circulation is restricted,many countries take steps to arrange for an official registration of these documents and their centralized signalisation.

Dissertations are research works submitted by those applying for a degree in science. Depending on their discipline and subject, dissertations may belong to different types of research work. In experimental works, the greatest attention is paid to the description of experiments carried out and the conclusions drawn. Methodological studies evaluate the existing methods, identify their heretofore unnoticed features, adduce fresh arguments in favour or against this or that method, or develop perfectly novel research or production methodologies. Descriptive studies give documental description of unknown scientific facts and phenomena, reveal their essence and interrelations with other facts and phenomena, and advance theoretical hypotheses proceeding from theoretical generalisations. Computational and analytic studies lay the emphasis on a mathematical interpretation of processes and phenomena. Finally, historical and bibliographical studies investigate into the role and importance of the individual scientists, trends and schools in the development of science and technology. In practice most dissertations are studies of a mixed pattern, possessing most if not all of the above features.

Dissertations are normally not published and are available but in two or three copies. What is published are 100 to 105 copies of abstracts 15-20 pages long, written by the author for the preliminary communication of the principal propositions to the scientific community.

Information cards can be used to serve as the main form of supply of scientific, technical, industrial and economic information by the local enterprises and organizations to the central agencies, that is as the main source of primary information. Information cards communicate on the accomplishments of research and development institutes, design and development bureaux, institutions of higher learning, enterprises, construction projects etc. A card is made in one copy immediately at the corresponding establishment in conformity with a standard pattern and signed by the head of the establishment.

Preprints are offprints of a paper or other work typographically printed before its official publication in the current issue of a periodical and intended to be sent out to a limited number of interested persons. Many periodicals and serials make offprints of the particular papers along with the main run and deliver them to the authors who use to send them out to their colleagues. This is a time-honoured practice and it has paved the way for preprints. These have the advantage of being ahead of the official publication of science documents, sometimes by several months, and of providing for faster communication of scientific facts and ideas to specialists.

Deposited manuscripts are a special category of unpublished primary documents. Manuscripts of monographs, collections of papers, and separate papers that are of interest to a limited circle of specialists can be transferred, on decision of publishers, journal editors or research institutes to a central information agency for deposition. This means that the manuscripts will be announced in special bulletins or catalogues and that their copies will be available on request from the information agency.

Unpublished documents are gaining importance as a means of communicating the latest results of scientific research.

The patterns of the distribution of scientific publications

So much for the main types of scientific literature. Until quite recently, a descriptive approach of that kind was the only one applied in the study of the system of scientific publications. Bibliologists and bibliographers aspired towards a most detailed typology of scientific documents and publications and wanted to ascertain the part played by each type in scientific work. The development of informatics has ushered in a new stage in the study of science publications. Science revealed some of the general regularities that characterize the inner connections existing between science publications and the growth of science, and derived quantitative relations between the number of publications and the indicators of the growth of science.

It has been ascertained that there exists a common and regular pattern of distribution for the number of authors who publish a given number of papers during their lifetime, the number of journals annually publishing a given number of papers, and the number of publications containing a given number of references to other publications. "They follow the same type of distribution as that of millionaires and peasants in a highly capitalistic economy. A large share of wealth is in the hands of a very small number of extremely wealthy individuals, and a small residual share in the hands of the large number of minimal producers. Whether the exact form of the distribution is lognormal, exponential, a Zipf Law, or an inverse square has been a matter of much conjecture in each of the cases".(1)

This means that the majority of authors produce but one or two papers in their lifetime, while a small group of writers are very proliferous and publish tens or even hundreds of papers. Periodicals that cease publication are for the most part those which had put out but one or two issues, while a small number of them issued since long ago account for the lion's share of all papers published. About half of all published literature is due to such number of authors or journals that is equal to the square root of their total number. "In short if there are for example 30,000 journals alive in the world, or 1,000,000 publishing scientists in a country then a selected list of about 175 journals or 1,000 scientists would account for half of the bulk of the literature but probably for 70-80% of the important content".(2) This is a discovery of great consequence: it helps to overcome the awe inspired by the tremendous amount of literature and provides theoretical grounds for finding the publications of a genuine value among the innumerable multitude of others.

In this connection, the questions arise: what is the proportion of the articles dealing with a given problem to the total number of scientific periodicals; and what is the proportion of the publications of genuine value to their totality.

Bradford's Law of Scattering

The former question has been answered two decades ago by British information scientist S. Bradford. He established that, taking for the unity the totality of publications dealing with a subject, the specialized periodicals in the field concerned will contain but one-third of these publications. Another third of the relevant articles will be found in a much larger number of journals of a different profile but having a relation to the given subject. And, finally, the last third of publications will be scattered among a vast number of periodicals in which it would have been impossible to predict their appearance as these periodicals are either of a too general nature or absolutely alien to the subject in question. This appearance of publications belonging to a field of knowledge in journals bearing no relation to that field is explained by the unity of science, by the fact that along with the differentiation of knowledge reverse integrative processes are unfolding.

---

1. Price, D. de Solla. Regular patterns in the organization of science. - "Organon", 1965, No. 2, p.246.

2. Price, D. op.cit., p.247.

Bradford has thus formulated the Law of Scattering of publications that he discovered: "If scientific journals are arranged in order of decreasing productivity of articles on a given subject, they may be divided into a nucleus of periodicals more particularly devoted to the subject and several groups or zones containing the same number of articles as the nucleus, when the numbers of periodicals in the nucleus and succeeding zones will be as $1 : n : n^2 \ldots$"(1)

Another British information scientist, B. Vickery,(2) refined the Law of Scattering, by deriving the following relation:

$$T_x : T_{2x} : T_{3x} = 1 : n : n^2,$$

where T_x is the number of journals containing x articles on a given subject; T_{2x} is the number of journals containing 2x such articles; T_{3x} is the total number of journals containing 3x articles; and n is any number depending on the value to which x has been set equal.

The scattering of publications over scientific periodicals, as has become clear, is a particular instance of the distribution characterized by the abovementioned Zipf Law. The Zipf Law is an expression of a universal property inherent in very different languages. The property consists in the fact that the bulk of any text in any language is made up of a very small number of frequent words while the rest of the words are used only rarely.

The law of scattering of publications is of great practical importance. It implies that the coverage of all the publications on a subject cannot be ensured by scanning only the periodicals of the given or related profiles; for this it is necessary to look through practically all the scientific and technical periodicals. In addition, the law can be used to determine the number of periodicals sufficient to provide for the coverage of a specified percentage of publications in a subject field or discipline. Later we shall describe how this law is taken into account in organizing information activities.

Bibliographic coupling of documents by citations

To answer the second question, it is necessary to get a notion of one of the novel methods of the study of scientific publications - known as the method of bibliographic coupling. It has struck the information scientists that it is possible to turn to account the fact that scholars regularly make references to the works of their predecessors who have explored the same, similar or related subjects. This practice, first emerged a century ago, has been for the past fifty years an indispensable standard and criterion of the thoroughness of a scientific publication.

---

1. Bradford, S.C., Documentation. 2nd ed. London, 1953, p.154.

2. Vickery B.C., Bradford's law of scattering. - "Journal of Documentation", 1948, vol. 4, No. 3, p. 200.

The study of bibliographic references contained in most publications has shown that they form a dense network of associations between scientific publications. Grouping of documents that cite the same publications reveals internal relations between the remotest subjects long before these relations can be otherwise noticed and explored. This is achieved by a purely automatic procedure since the writers of publications on most diverse subjects may cite the same document that they regard as their point of departure.

Citations of the materials used have long been serving scientists and practicians as important clues in searching for the relevant information sources. This path, however, is long and laborious, and, what is worse, it does not ensure the detection of all the pertinent works. For a better insight into the system of citations, let us consider some single subject of research.

American biologist G. Allen compiled a map of the citations of the earlier documents in fifteen articles on nucleic acids published between 1941 and 1960. This map, represented in Fig. 4, shows that for the complete coverage of all the 15 articles on the basis of citations they contain it is necessary to refer to six articles at least: 6, 7, 9, 13, 14, 5, published between 1956 and 1958. Some articles that are rarely or never cited-namely, 7,8, and 9 - are hard to find by this technique. It is quite natural, too, that the latest publications (14 and 15), issued simultaneously, are not connected by reciprocal citations.

If it were possible to trace the bibliographic relations between publications on the basis of citation data not from bottom to top, that is from the recent works to the earlier ones, as has been done usually, but vice versa, from top to bottom, that is from earlier to the recent documents, this would make the finding of all the articles pertaining to a given subject much easier. References to article 2 alone would identify at once a half of the remaining articles - 3,5,6,7,11,12,15, while references to three articles 1,2,4, would provide information on almost all the publications, leaving out of sight but two of them - 9 and 13. Of late, this has become possible. The bibliographic coupling method has found its practical implementation in citation indexes compiled with the aid of electronic digital computers. These will be dealt with later on. Meanwhile we are going to dwell on the application of this method for a theoretical study of the system of science publications.

It might be asked, to what extent bibliographic coupling of documents provides for an adequate and exhaustive reflection of the connections actually existing in science. To answer this question, E. Garfield, the author of the first modern citation index and head of the Institute for Scientific Information (Philadelphia), staged the following experiment in 1964.(1) The history of a most important discovery of the present-day biology - the mechanism of transfer of hereditary characteristics through desoxyribonucleic acids - was subjected to parallel studies by means of conventional techniques and by the bibliographic coupling method. The conventional method was based on a historical monograph by

---

1. Garfield, E.a.o. The use of citation data in writing the history of science. Philadelphia, 1964. V, 75 p.

the well-known American scholar and writer Isaac Asimov entitled "The Genetic Code", while the bibliographic coupling used a special computer-produced citation index to genetics literature.

The comparison showed that 65 per cent of the publications identified by both methods as turning points on the way towards the discovery were identical. The conventional method proved more efficient only in identifying the publications of the last century when the practice of making references to the papers used was insufficiently developed. As to the past few decades, bibliographic coupling spotted a great many associations between publications missed in Asimov's book. The experiment granted a number of untrivial conclusions. The most remarkable one was that bibliographic coupling of documents on the basis of citation data does provide for a correct and full recognition of the actual relationships between research papers. The assumption was corroborated, too, that the greater the attention paid by the scholars to some publication - and consequently the greater the number of references they make to it - the greater its scientific worth.

We have already mentioned that the distribution of citations over the citing articles follows the general pattern of distribution of mass-scale phenomena, it is analogous to the numbers of authors and journals publishing a given number of articles during their lifetime. To illustrate this by figures: 10 per cent of scientific publications make no references at all, 85 per cent of publications containing from 1 to 25 references account for a half of the total citations, while the remaining 5 per cent of publications with a very great number of references - surveys and reviews - account for the other half. Despite the fact that some publications lack references altogether, while some may contain a sizable number of references to earlier works, certain average figures can be derived. These are: 15 citations per publication, of which 12 or 13 are references to journal articles.

At this point it is possible to try to derive the ratio of the publications of genuine worth to their total number. For this we shall observe how the new works put out in a field during one year are linked by citations to all the publications existing in that field. /Fig.5/. To make the picture clearer suppose that by the beginning of the year there were 100 publications dealing with the field or subject under consideration. Assuming the regular exponential growth of literature (a 7-per-cent annual increment), seven new papers will be published during the year, each containing - according to the above average data - 13 references to earlier journal articles. The total for the whole year will be 91 references (13 x 7) to earlier publications.

As shown by a special analysis, the new citations will have the following pattern of distribution with respect to the earlier publications: 40 of the earlier papers will not be mentioned at all, 10 references will be to publications in other fields of knowledge, 50 will be to 50 papers in the field, that is each of these will be mentioned once during the year, and 31 references will be to 10 papers cited more than once. These latter 10 papers (their associations with the 7 new ones are shown on Fig. 5) are the most important for science. (Recall that the more often a paper is cited in later publications the greater its scientific value.)

We see that the new publications tie together, as it were, by their references but a small part of the previous literature, which may be regarded as the active front of research, while the rest of the literature is on the whole of minor significance.

It is interesting to analyse the dynamic characteristics of this active research front and assess its relative distance from the publications of the preceding year. For this purpose, 200 consecutively issued papers constituting the literature of a narrow topic in physics were studied. In the autumn of 1903, French Academician R. Blondleau announced his discovery of a new radiation that he called N-rays. This communication gave rise to wide discussions and many publications until at last the discovery was recognized to be a mistake and the N-rays mere fancy of the famous physicist. The episode, well known to the historians of science, is convenient for building a model of the distribution of citations over publications as it is definitely localized and fixed in time.

Figure 6 shows a matrix where every dot marks a citation of a paper in another one. The horizontal axis shows the numbers designating the citing papers and the vertical axis, the cited papers, designated by the same numbers. The papers are numbered chronologically, as published. The matrix is a graphic illustration that a great number of references fall in the band of the 50 papers immediately preceding each publication concerned. Thus the continuous and the dashed lines are the boundaries between which lies the active research front identified on the preceding diagram.

Remarkable are the vertical bands of dots corresponding to the publications that contain the maximum number of references and the horizontal bands that correspond to the most frequently cited papers. The former bands indicate that after every 30 or 40 publications the need arises for a survey that would serve as substitute for the older publications which are lost out of sight behind the active front. The latter bands point to the classic works in the given field which are perpetually referred to with a roughly constant frequency and constitute but a minor percentage of the total publications.

This model can be considered in general as showing the pattern of the distribution of citations in publications dealing with any subject, problem or discipline, and hence, as a characterization of the distribution of the most frequently used publications over their totality.

The regularities of the obsolescence of scientific publications

Studies of bibliographic references relatively long ago made it possible to establish certain regularities of the obsolescence of science publications. Thirteen years back American bibliographer Ch. Brown analysed some 40,000 citations in the leading journals published in English, French and German in mathematical, physical, chemical, geological and biological disciplines in the period 1952-1954. He took several volumes from each of the 56 journals and serials covered in his study to a total of 40,000 pages. Among the many conclusions he made we are now interested in his data on the obsolescence of scientific books and periodicals. These show that the faster the growth rate of a scientific discipline the quicker its literature becomes obsolete. In the journals covered in Brown's study, references to books dated at most ten years back accounted for 64% of all references to books of physics and chemistry, 62% of physiology, 56% of geology, 38% of botany, and 27% of mathematics. The picture was much the same with the references to journal articles. However, the percentage of references to papers younger than ten years was conspicuously higher than that to books, and the obsolescence rate of the papers of preceding decade was more rapid.

To characterize the process of obsolescence of literature in a field of science, a concept borrowed from nuclear physics has been applied of late. By analogy with the life period of disintegration of radioactive substances an attempt was made to measure the periods of half-life of scientific literature. Half-life is defined as the time during which a half of all currently utilized literature in a field has been published. It is exactly equal to the time during which a half of all the literature published in the field by the present moment will pass out of use. This fact is of major importance as it enables us to make reliable predictions of the future relative worth of publications on the basis of an analysis of the current obsolescence rates of the literature in the field concerned.

Proceeding from the findings of Brown (supplemented by their own data) librarian R. Burton and physicist R. Kebler of the United States computed the following half-life periods of papers in the different branches of science:

Physics 4.6 years
Physiology 7.2 years
Chemistry 8.1 years
Botany 10.0 years
Mathematics 10.5 years
Geology 11.8 years

This shows that a great part of physical papers becomes obsolete very soon, being so to say ephemeral literature, that in physiology and chemistry there are approximately as much short-lived as there are lasting, classic works, while in mathematics the latter type is predominant. The age of publications that constitute the active research front depends, it will be recalled, on the field of knowledge. It is shorter in the fields characterized by shorter half-life, such as physics, physiology, and chemistry, and longer in the fields with longer half-life, such as botany, mathematics, and biology.

We are now able to derive the general averages, too. Fig. 8 shows the age distribution of works cited in the publications of 1965. The data are based on an analysis of 3.3 million scientific papers issued within several hundred years and cited in 296,000 publications put out in 1965 in the exact, natural, and applied sciences.

The greatest proportion of the cited papers (13 per cent of the total) are dated 1964. Since fresh publications have to wait before the scientists get aware of them, the percentage of references to the publications of 1965 is naturally much lower (5 per cent of the total citations made that year). Then follows a decline the middle of which (6.5 per cent of publications) falls to 1960. Publications dated from 1960 to 1965 make 52.6 per cent of the total citations made in 1965. This means that the average lifetime of the publications constituting the active research front does not surpass 5 years.

To sum up. We have seen that among the many types of scientific literature, this major means of knowledge dissemination, of paramount - though varying - importance are journal articles, books, and scientific and technical reports. Recent research has shown that far from all of the huge number of scientific documents accumulated by mankind possess any real significance for the further research. The analysis of the interrelations between science publications on the basis of bibliographic citation data helped to establish that the core of the scientific works of genuine value is made up of but a minor proportion of the publications of recent years, namely those most frequently cited in later books and papers. It is these works which constitute the active research front, that should be sought first of all.

All the subsequent lectures will deal with the different aspects of the available means and methods for the searching of these publications and their efficient utilization.

Checkup Questions

1. What are the development stages through which written records have passed from antiquity till our day?
2. What is a scientific document and what are the existing types of scientific documents?
3. What is book and what kinds of books are important for scientific work?
4. What are the existing types of periodical publications and what are their characteristic features?
5. What does the importance of journals and journal articles lie in?
6. What functions are fulfilled by the special types of technical publications?
7. What are the contents of scientific and technical reports?
8. How are scientific publications interconnected by bibliographic references?
9. What regular patterns of distribution of scientific publications do you know?
10. What regularities of the obsolescence of scientific publications do you know?

Literature

1. Johnson, E.D. Communication. A concise introd. to the history of the alphabet, writing, printing, books and libraries. New Brunswick (N.J.), Scarecrow, 1955. 211 p.
2. Kronick,D.A. A history of scientific and technical periodicals: the origins and development of the scientific and technological press, 1665-1790. New York, Scarecrow, 1962. 274 p.
3. Price, D.J. de Solla. Networks of scientific papers. - "Science", 1965, vol. 149, 3683, p.510-515.
4. Burton, R.E. and Kebler, R.W. The "half-life" of some scientific and technical literatures. - "American Documentation", 1960, vol. 11, 1, p. 18-22.
5. Merton, R.K. The Matthew effect in science. - "Science", 1968, vol. 159, 3810, p. 56-63.

3. INFORMATION AND BIBLIOGRAPHIC PUBLICATIONS - SOURCES OF DATA ON LITERATURE

As was established in the previous lecture, scientists at all times considered scientific literature to be too abundant and diversified to serve as a medium for dissemination of scientific knowledge. Therefore, along with those kinds of scientific literature which we have studied, special kinds of scientific documents were evolving whose primary mission was to serve as guides to the literature of science and to scientific documents.

Stages in the development of secondary documents

Even hoary antiquity saw the existence of lists of scientific books. Among the clay tablets dating back to the 2nd millenium B.C., there were found some on which no texts and only titles were inscribed. They were prototypes of present-day book catalogues. In ancient Greece and Rome very large and highly sophisticated book lists were compiled. The "Tablets" of the scholar of antiquity Callimachus, compiled around 250 B.C., comprised 120 volumes and contained detailed and accurate data on the author, title and size of books, and the time when they were written. These lists or catalogues gave rise to a special kind of literature - bibliography. About the close of the 4th Century St. Jerome wrote "De Viris Illustribus," which is believed to be the earliest of all known bibliographies.

A century and a half following production of the first printed books, scientific periodicals made their appearance; as you already know, they have for a long time served to announce to scientists new scientific publications. Another hundred and fifty years later, in mid-18th Century, Diderot's and D'Alembert's "Encyclopedie, ou Dictionnaire universel des arts et des sciences" started publication, giving a powerful spur to the development of reference literature. The first quarter of the 19th Century saw the beginning of national (State) bibliographic registration of printed materials regular publication of national bibliographies, and the appearance of first abstract journals interpreting the contents of journal articles.

Each of these document types performs some useful function within the contemporary system of scientific literature, and their appearance was in response to a certain social need, as often as not unrealized at the time, which was surely generated by the growth in the number and variety of scientific literature. Therefore, although we often express our dissatisfaction with the historically evolved system of communication of scientific knowledge, we cannot deny that its progress is subject to certain regularities.

According to the previously established typology of scientific documents, those which contain mainly information on scientific literature or results of its analytico-synthetic processing are placed among the secondary documents. As the majority of these are published in the course of information or bibliographical work, we call them information and bibliographic publications.

Types of secondary documents

If the degree of information processing be taken as a criterion, the different kinds of secondary publications can be conveniently placed in the following order: reference literature, reviews, abstracts and annotations, bibliographies, catalogues and card indexes. Here the first document kinds are those in which information has been subjected to maximum processing, while the last ones are those where it was minimal. We shall adhere to this order in characterizing secondary documents and publications, which are the subject of this lecture. We shall speak only of the basic types and give the respective illustrating examples. The particular titles of information and bibliographic publications which are necessary for searching scientific literature will be cited in selective annotated lists printed in the Appendix.

Reference literature is intended for rapidly obtaining data of scientific, applied or instructional character. These data include results of theoretical generalizations, basic scientific facts, mathematical, physical and chemical constants, industrial information accompanied by various tables, diagrams, drawings and formulas. The most significant of reference publications are encyclopaedias, industrial handbooks, glossaries, vocabularies of terms, biographical dictionaries, bilingual and multilingual dictionaries.

Encyclopaedias are reference publications which contain the most essential information (comprehensive or brief) on every or some particular field of knowledge and practical activity. They divide, according to content, into general (universal) and those devoted to a particular topic; according to structure, into alphabetical and systematic; according to size, into desk-type (1 to 4 volumes), small (5-12 volumes) and great (several dozen volumes). Encyclopaedias are usually written on the basis of a topical plan, by which the bulk of the work is broken down into different scientific disciplines or their parts and, of a word-list giving all the words for which separate entries are prepared.

Entries in encyclopaedias vary in type and length. The largest are review entries covering broad topics. The majority of entries are references which provide a definition and some basic information on a subject. Short interpretative entries include only a definition and sometimes the origin of the word. Finally, encyclopaedias contain a great deal of cross-reference entries which indicate that the relevant information should be looked up under a different word. A great help are the bibliographical lists appended to entries: they enable the reader to study at greater length a subject only superficially treated in the encyclopaedia. Many encyclopaedias are provided with subject indexes which heighten their value as reference tools.

General encyclopaedias are represented in the U.S.S.R. by all the three types: great, small and desk-type. The Great Soviet Encyclopaedia (GSE) has appeared in two editions thus far. The first, completed in 1947, comprises 65 main and one supplementary topical volume "U.S.S.R." The second appeared in 1949 through 1958; it comprises 49 main and 2 supplementary volumes. New data are published regularly in GSE Yearbooks. There is a two-volume alphabetical subject index. The Small Soviet Encyclopaedia has passed through three printings: the first edition, in 10 volumes, was brought out from 1928 through 1931; the second, in 11 volumes, from 1933 through 1947, and the third, in 10 volumes, from 1958 through 1961. For the latest edition, containing some 48,000 entries, a 165,000-word subject and name index was printed in a separate volume. Desk-type encyclopaedias are represented in the U.S.S.R. by the Encyclopaedic Dictionary which is intended for the general public. It was brought out in three volumes in 1953-1955, and in two volumes, in 1963-1964.

Of the general encyclopaedias, the oldest and probably best known is the Encyclopaedia Britannica, or Britannica, which has been appearing for nearly two centuries past. Its first three-volume edition went off the press in Edinburgh in 1768-1771. The Britannica is traditionally considered to be one of the most authoritative encyclopaedias. This is particularly true of the second, 29-volume, edition published in 1910-1911. It was that edition to which the eminent Russian scientist K.A. Timiryazev referred as the first genuine encyclopaedia of our century. A feature of the Britannica is its small word-list and most minute review entries; it contains some 43,000 entries, with the subject index running up to 500,000 entries. The latest completely renewed edition of the Britannica was the 14th edition in 24 volumes, brought out in 1929.(1) It has been reprinted annually since then, with one-eighth of the entries undergoing changes. The encyclopaedia thus renews its contents rather rarely and contains much outdated information. Besides, this type of encyclopaedia is no longer wholly satisfactory to the scholar; with the ever deepening differentiation of scientific knowledge the system of review-type entries detracts from the value of the publication as a reference tool.

In sharp contrast to the Britannica stands the multi-volume German encyclopaedia Der Grosse Brockhaus, which was first published in 1796-1811. It contains a great many reference-entries and is intended for the broad reading public. Its latest, 16th edition comprises 12 main and 2 supplementary volumes totalling more than 150,000 entries.(2) Der Grosse Brockhaus is distinguished by a uniform coverage of the whole universe of knowledge, with a slight bias towards natural sciences. It can be said to be the most up-to-date, detailed and carefully prepared of all existing encyclopaedias.

Among French encyclopaedias, the greatest popularity is enjoyed by the numerous Larousse publications. Natural and exact sciences and technology feature rather prominently in the 10-volume Grand Larousse encyclopédique, which provides the most up-to-date and detailed information on nuclear physics, astronautics, electronics and similar disciplines.(3) Particularly widely used is the Petit Larousse, first published in 1966 and reissued annually ever since.(4)

---

1. Encyclopaedia Britannica. 14th ed. Vol. 1-24. Chicago-London, 1929.
2. Der Grosse Brockhaus. 16 Aufl. Bd. 1-12, Ergänzungsband, Atlasband. Wiesbaden, 1952-1958.
3. Grand Larousse encyclopédique. Vol. 1-10, Paris, 1959-1964.
4. Petit Larousse. Paris, 1967.

Encyclopaedias for special topics differ from general encyclopaedias not only in coverage but in a more limited purpose as well, being, as a rule, designed for specialists and researchers. To cite a few examples, the following publications were brought out or reissued in recent years:

- McGraw-Hill Encyclopaedia of Science and Technology. Vol. 1-15, Yearbooks. New York, 1960-1969.
- International Dictionary of Applied Mathematics. New-York, Van Nostrand, 1960.
- Kirk R.E. and Othmer D.F. Encyclopaedia of Chemical Technology. New ed. Vol. 1-18, Supplements. New-York, Interscience, 1963-1969.
- Gray P. (ed.) The Encyclopaedia of the Biological Sciences. New York, Reinhold, 1960.

Encyclopaedias can serve as sources of scientific information. However, a critical approach is warranted in using an encyclopaedia. The latest editions are to be preferred, and the date of publication printed on the title-page should be checked in order to verify whether it stands for a new edition or merely a reprint. The rule of thumb to follow with respect to encyclopaedias is summarized in the words of the distinguished historian of science G. Sarton who said that to turn to an encyclopaedia for a primary reference is clever, to ignore encyclopaedias is unreasonable, to rely on them too much is stupid.

Handbooks provide data which can be useful for different subject specialists. They vary broadly in subject-matter, structure, and size. For example, there are very compact handbooks, such as:

- American Institute of Physics. Handbook. New York, McGraw-Hill, 1957.
- Lange N.A. (ed.) Handbook of Chemistry. 10th ed. New York, McGraw-Hill, 1961.
- Machinery's Handbook: A reference book for the mechanical engineer, draftsman, toolmaker, and machinist, 17th ed. New York, Industrial Press, 1964.
- Flugge W. (ed.) Handbook of Engineering Mechanics. New York, McGraw-Hill, 1962.

Such a classic work as "Beilsteins Handbuch der organischen Chemie"(1)comprises about 100 volumes, which have been brought out in three cycles from 1918 to this day. The same great size and elaborate structure are typical of the well-known

---

1. Beilstein's Handbuch der organischen Chemie. 4. Aufl. Bd. 1-27. Berlin, 1918-1940, Bd. 1-31. 1928-1938, Bd. 1-27. 1941-1955, Bd. 1-. 1963-.

inorganic chemistry handbook popularly known as "Gmelin".(1) The hundreds of its issues and their indexes call for a special study before a specialist can make full use of them.

Closely associated with handbooks are tables of constants, guides to minerals, plants and animals, drawing and map atlases, collections of formulas and diagrams, and similar publications, for example, Handbook of Mathematical Tables. Supplement to Handbook of Chemistry and Physics. Cleveland, Chemical Rubber Co., 1962; Kaye G.W. and Laby T.H. Tables of Physical and Chemical Constants and some Mathematical Functions. 12th ed. London, Longmans, 1959.

Biographical (who's who) dictionaries supply data on the life, activities and works of eminent scientists, engineers, physicians, agronomists, public figures etc. Such dictionaries are particularly useful in studying the history of a science, technology and production. The examples are:

- American Men of Science. Vol. 1-5. Arizona, Cattell, 1960.
- Directory of British Scientists. London, Benn, 1963.
- World Directory of Mathematicians. Bombay, Tata Institute of Fundamental Research, 1961.

Dictionaries may be a very useful starting point as an aid to establishing a clear definition and the precise meaning of a term. Any single-language dictionary is concerned essentially with information about word-spelling, pronunciation, derivation, meaning and usage, e.g. The Concise Oxford Dictionary of Current English. Ed. by H.W. Fowler and F.G. Fowler. 4th ed. Rev. by E. McIntosh. London, Oxford Univ. Press, 1956.

Translating (bilingual and multilingual) dictionaries contain words and terms of one language and their equivalents in other languages. They can be either general or special, i.e. explaining terms in a special subject field, for examples:

- Hoyer-Kreuter's Technological Dictionary in three Languages: English-French-German. New York, Ungar, 1944.
- Callaham L.I. Russian-English Chemical and Polytechnical Dictionary. 2nd ed. New York, Wiley, 1962.
- Emin I. Russian-English Physics Dictionary. New York, Wiley, 1962.
- Russian-English Atomic Dictionary. New York, Technical Dictionaries, 1957.

---

1. Gmelin-Krauts Handbuch der anorganischen Chemie. 8. Aufl. Klausthal-Zellerfeld, 1924-1965.

Guides to general reference books contain bibliographical details of such books. The most important guides are:

- Walford A.J., Guide to Reference Materials. 2nd ed. London, Library Assoc., 1966-1969.
- Winchell C.M., Guide to Reference Books. 7th ed. Supplements 1-4. Chicago, American Library Association, 1951-1963.

We can mention also: Maichel K., Guide to Russian Reference Books. V.1. General bibliographies and reference books. Stanford, Hoover Institution, 1962. 92 p.

Reviews of progress in different fields of science and technology have been in existence for quite a long time. In the last decade they have come to occupy a steady position among the other secondary publications. Surveys summarize the data contained in primary documents, such as articles, monographs, conference reports, technical reports, invention specifications. They generally cover materials in one subject field over a definite time period. Surveys can be analytical or informative.

Analytical reviews are distinguished by deeper analysis and based on appropriately ordered data selected from primary scientific publications. They usually contain all basic facts concerning the progress in a subject field, and permit to assess the advances in the study of an issue or problem or in an area of research.

Surveys are very valuable sources of information. They initiate an incipient researcher into his field of interest, familiarize him with the progress in related fields of science and technology, help to choose or more precisely define the direction and methodology of his own study or design work; they are great time-savers, sparing a specialist long hours of searching and scanning of primary materials.

Surveys are published in different forms, such as an article in a common journal, or in a special reviewing journal (e.g. "Advances in physics" , a separate issue of a periodical, or of an irregularly appearing book series. They are typified by surveys of different disciplines published by the British quarterly "Science progress", which prints review papers on advances in astronomy, meteorology, physics, chemistry, geology, mineralogy, zoology and botany. Among others, the following surveys have been published in recent years:

- Annual Report of Progress in Physics (Inst. of Physics and Physical Society).
- Chemical Reviews (American Chemical Society).
- Advances in Space Science. Vol. 1-11 New York, Academic Press, 1959-1969.
- Progress in Aeronautical Sciences. Vol.1-9. London, Pergamon, 1961-1969.

Abstract journals are probably the most important secondary information publications now. They appear regularly and reflect the contents of the most recent scientific documents in abstracts or, to a lesser extent, in annotations.

An abstract is a summary of a document specifying its character, as well as the techniques, findings, time and place of an investigation. An abstract also indicates the objectives and main theoretical premises of a study, supplies the pertinent figures, formulae, tables, charts and diagrams. In conclusion, the number of illustrations, tables and references in the abstracted publication are indicated, as well as the abstractor's name. All abstracts fall into two general types: informative (abstract-synopsis) and indicative (abstract-summary). An informative abstract contains in an abridged form all the main ideas of a primary publication, while an indicative abstract recognizes only the most important of them. An abstract, as a rule, does not reflect the abstractor's subjective views, and does not make any judgments on a document.

It is worthwhile to cite here the Guide to the preparation of abstracts formulated by one of the leading abstracting services, "Biological Abstracts": "An abstract should be a non-critical, informative digest of the significant content and conclusions of the paper, not a mere description. It should be intelligible in itself, without reference to the original, but is not intended to substitute for it. It should be brief (preferably less than 3% of the original), written in whole sentences, not telegraphic phrases".(1)

Naturally, the size of an abstract depends on the size and importance of the abstracted document. However, the abstract length in different abstracting services has stabilized at the 800-900 character mark. The average length of abstract in some leading abstract journals in 1965 was as follows: "Physical Abstracts", 500-550 characters; "Chemical Abstracts", 900-1000 characters; VINITI's Referativnij Zhurnal, 1100-1200 characters.

Annotations, in contrast to abstracts, are not summaries of primary documents but only short communications on their theme or content, although they also detail their character. According to purpose, annotations divide into descriptive annotations, which provide a factual description of the content and some other details of a document, and recommendatory (critical) annotations, which give an appraisal of a document from the point of view of its usefulness to a particular user category.

Information and bibliographical services usually prefer the descriptive type. These may be very short, consisting of a few words or short phrases, or more detailed, but in either case they only characterize and not expound the main ideas and conclusions of a primary document. This feature makes annotations particularly suitable for revealing the contents of books. The reason is that the multidimensional content of books does not lend itself easily to abstracting. A descriptive annotation usually details the kind of book, its purpose, and structure, the scientific school to which the author belongs, the topic of the

---

1. Biological Abstracts, 1969, vol. 50, 1. p. XXI.

study and its conclusions, as well as the auxiliary and illustrative material, supplements, addenda, and the reference apparatus, including indexes and bibliography. An annotation does not on the average exceeds 400 to 600 characters. A characteristic feature of the annotation is that it must be closely associated with the description contained in the entry, which in abstract journals immediately precedes each abstract and annotation, but may also be included as a separate entry, that is as a secondary document in its own right. In addition, each document indexed in an abstract journal is assigned a registration (serial) number and a classification number (document symbol). For foreign materials, a translation of the title is also given and the language of the original document is indicated.

Abstract journals have two main objectives: to keep specialists in touch with all significant current developments in their respective fields, without recourse to primary publications, and to help them find materials on a specific subject published over a fairly long period of time. The first objective is taken care of by the publication of monthly issues of specialized series of an abstract journal, in which the material is grouped by subjects, and by their distribution on a subscription basis among specialists and research libraries. The second objective is ensured by the regular printing of subject, author, formula and other special indexes to each issue or yearly set of the journal.

It has to be admitted that abstract journals are not up to the mark with respect to either of these objectives: in the former case, because of considerable publication delays (an abstract appears in an abstract journal at best four months after publication of the original paper); in the latter case, because of the inadequacies of reference tools and lack of cumulations (in searching for literature over a fairly long period of time, a large number of annual indexes have to be looked up, and a portion of indexed materials will always escape the search). However, abstract journals are steadily improving their quality and also grow in number; at present they run up to several hundred titles.

In using different abstract publications, the quality, type and size of abstracts, their arrangement, the time lag between the appearance of a publication and of its abstract, the availability and quality of indexes, are not the only features to be considered. An equally important factor is the coverage of scientific publications, their languages and country of origin, as well as their categories. In other words, it should be known whether papers alone or also books, patent specifications, technical reports are abstracted, and whether these are only in European or in Oriental languages as well, etc. By way of illustration, we shall present data on several well-known abstract journals.

Referativnij Zhurnal, the Abstract Journal of the All-Union Institute of Scientific and Technical Information (VINITI) has been published since 1952; in 15 years it has grown to be a unique service, the largest of its kind in the world.

The Referativnij Zhurnal comprises the following series: Automation and Radioelectronics, Astronomy and Geodesy, Biology, Biological Chemistry, Geography, Geology, Mining, Geophysics, Mathematics, Mechanical Engineering, Metallurgy, Mechanics, Physics, Chemistry, Industrial Economy, Electrical Engineering. In

1963, the separate series "Scientific and Technical Information" was started to cover materials in the discipline which is the subject of our lectures (from 1966 it appears also in English). Each series comprises annually 12 issues included in cumulative volumes, with the "Chemistry" series comprising 24 issues. The material is arranged systematically, i.e. according to the framework within which different problems are treated in the respective areas of science and technology. This framework is presented in a special publication, "Subject Headings of VINITI Abstract Journal", which is published annually. Some large portions of the main series included in the cumulative volumes appear in separate series of issues for the convenience of narrow subject specialists.

The Abstract Journal surpasses all similar publications in the length of its abstracts which average 1100-1200 characters and approximate the informative type. The time lag between the appearance of a publication and of its abstract has in many instances been reduced to 4 months, which is quite a low figure. Each series is provided with indexes: an author index to each issue and an annual one, and an annual subject index; the "Chemistry" series has, in addition, annual classified, formula and patent indexes. The VINITI Abstract Journal covers articles and materials from over 20,000 periodicals in 70 languages printed in more than 100 countries. The most significant scientific books and patent specifications are also abstracted.

The "Chemical Abstracts", brought out by the American Chemical Society since 1907, covers scientific and technical materials in theoretical and applied chemistry, chemical technology, metallurgy and metallography, chemistry of foodstuffs, fuels, textile and leather industries, etc. Its abstracts are somewhat shorter than those in the VINITI Abstract Journal, being an average of 900 to 1000 characters in length. The material is arranged systematically. The journal appears semi-monthly, resulting in two volumes a year comprising 13 issues each. Each issue is provided with an author and subject indexes; each volume, has an author, subject and formula index; there are also cumulative indexes for 10 years (the latest appeared in 1956), and a list of abstracted periodicals including 7,000 titles. The distribution of the periodicals by countries is approximately as follows: USA, 25%; USSR, 20%; Britain, 15%; Japan, 10%; Germany (East and West), 10%; France, 5%; and other countries, 15%. The weekly current awareness service "Chemical Titles" is produced by means of a computer, and covers 350 (mainly English-language) chemical publications. Fig. 9 compares two abstracts of the same paper which appeared in VINITI Abstract Journal "Chemistry" series and in "Chemical Abstracts".

Among abstracting publications the following can be particularly useful: Engineering Index, United States Government Research Reports, Mathematical Reviews, Science Abstracts, Nuclear Science Abstracts, Analytical Abstracts, Biological Abstracts, Building Science Abstracts, Computer Abstracts, International Aerospace Abstracts, Metallurgical Abstracts, Copper Abstracts, Zinc Abstracts.

Express-information service is a broad and rather vague term embracing a variety of information services, e.g. sets of printed book cards, lists of new articles, lists of current titles of periodicals. The only feature they have in common is that all of them were established to speed up information on new publications. The emergence of this type of service is directly related to the considerable delays in most of the existing abstracting services, which is reduced multifold in express-information publications.

Recently, the term "express-information" has came to be increasingly applied also to periodical bulletins containing detailed abstracts or digests of the most significant inventions, discoveries or investigations, which must be promptly made known to specialists. This kind of publication is typified by VINITI's Express-Information which appears (in Russian) in 50 series on problems of topical interest, e.g. "Astronautics and Missile Dynamics", "Computer Technology", "Automatic Control of Production Processes", "Synthetic High-Polymer Materials", or on major branches of economy and industry, e.g. "Radio Engineering and Electronics", "Oil Production", "Railway Transport", "Ferrous Metal Industry". A total of 48 issues each comprising 5 to 8 digests of articles, reports, invention specifications, i.e. some 300 different papers, are printed yearly in each series. Digests in the VINITI Express-Information appear approximately in a month's time after the original publication.

Library catalogues are listings of books held in a library, designed for the benefit of the readers to serve as a key to the library's stock. Library catalogues are the principal means to reveal the contents of library collections, to popularize literature and to guide reader in self-education. Catalogues have much in common with bibliographical card indexes, both being listings of publications. The distinguishing feature of the catalogue is that it is confined to the contents of a particular library, whereas a card index includes materials irrespective of their locations. Catalogues differ in the arrangement of entries, in the categories of material included, in the coverage of the library holdings, in the language of the material covered, in their purpose, in physical form, and some other characteristics.

According to the arrangement of entries, catalogues are divided into classified, alphabetical subject, and author and name catalogues.

In order to facilitate the selection of literature by subject, and to meet the thematic requests of readers, information agencies and libraries keep classified and alphabetical subject catalogues. A classified catalogue arranges subject entries for books according to the parts of knowledge with which these books are associated. Within a subject field the entries are first grouped in classes, then in subclasses, and so on. The structure of a classified catalogue is directly related to the library or bibliographical classification adopted by the library. Some of these classification schemes will be introduced to you later. Classified catalogues serve as a useful tool in scientific information and play an important role in library activities. Grouping library materials by fields of knowledge greatly facilitates reading guidance and selection of books. Classified catalogues are also used for in-house operations, such as acquisition or bibliographic and reference work.

From what has been said above one may be led to believe that a classified catalogue will always be instrumental in establishing what materials on a particular subject are available in a reference or library stock. However, if a subject is quite specialized, and the reader is not an expert in the given subject field, he will be hard put to it to determine under what subdivision in the classification scheme used in the catalogue the question of his interest should be looked up. To facilitate work with a classified catalogue, alphabetical subject indexes, also called "keys", are constructed, which list in an alphabetical order all the subjects contained in the catalogue together with their symbols (classification numbers). An index to a classification schedule may also include subject headings which provide different formulations or groupings of the topics reflected in the classification, as well as synonyms of terms appearing in the schedules.

But even the key to a classified catalogue will sometimes fail when all the materials on a particular topic are to be found. Imagine yourself searching for the most recent publications pertaining to aviation. In a classified catalogue they will be scattered over the different parts of knowledge, according to the aspect in which a given topic is presented, e.g., aircraft design - in engineering, aircraft construction - in industry, aircraft operation - in transport, military aircraft - in military science, occupational diseases connected with flying - in medicine, etc. Similar situations arise when one is looking for documents treating any other composite problem, such as a disease (e.g., problem of cancer), manufacturing process, farm crop, literature pertaining to a particular locality, or to the life and work of a person who contributed greatly to science, technology or public affairs. Often in such cases a catalogue which groups literature not by field of knowledge (as a classified one) but by some other, subject complexes, may prove to be helpful.

Alphabetical subject catalogue also reveals the contents of documents, but the entries are filed alphabetically by the subjects dealt with in the documents. The word or phrase denoting any one subject (notion) and used to express the subject-matter of a document or of its part is called a subject heading. The arrangement of subject entries in an alphabetical subject catalogue will be made clearer by drawing an analogy between it and the already familiar book types, notably, a textbook where the material is arranged systematically, and an encyclopaedia where the material is arranged by subjects. Or take any scientific book: there the table of contents will simulate a classified catalogue, while the index will stand for an alphabetical subject catalogue. Alphabetical subject catalogues are eminently suitable for reflecting publications in applied sciences--technical, agricultural or medical. This is the principal reason for their popularity chiefly with special libraries, which use them to cover both domestic and foreign materials, books and periodical articles.

The author and name catalogue lists works, irrespective of their content, alphabetically under their authors, names of institutions, or, in case of anonymous works or works having more than three authors, under titles. Such an arrangement enables the author and name catalogue to answer two principle questions: (1) Has the library a specific work whose author and title are known? and (2) What works of a given author, which have appeared as separate published items, are available in the library? If an author and name catalogue includes a sufficient number of added entries under persons that took part in the preparation of certain books, it will provide answers to some secondary questions (e.g. in what works did a given person participate as compiler, editor, illustrator, translator, etc.). The importance of the author catalogue is determined by it providing a facility of answering questions on the availability of this or that item in library stock. The author catalogue is thus one of the principal reference tools of an information agency or a library. This is explained by the fact that most readers when they come to a library already know what books they want. They use the author catalogue to check on the books sought, and find the class marks by which they are filed in the stacks. All library operations - acquisition, processing and storage of the holdings, service to readers, bibliographic and reference work and social activities - are only made possible by the existence of an author catalogue. Besides, an author catalogue is the only means of reflecting what authors are represented in library holdings. It provides for the popularization of the works of domestic and foreign classics of science, of outstanding personalities of today in all walks of social life. Therefore it would be wrong in all walks of social life. Therefore it would be wrong to treat the author catalogue as a formal one and to allot to it a subordinate place among the other types of catalogue.

The catalogue types that have been described are the principal ones as regards the method of arrangement of their entries. Some documents, e.g. standards and patent specifications, are more usually sought by their numbers than by titles. Therefore, numerical catalogues are used in which entries are ordered not alphabetically by titles, but numerically by the numbers assigned to them. Combinations of the principal types of catalogue are occasionally encountered. For instance, some special libraries maintain alphabetico-classed catalogues in which the main divisions are arranged systematically, i.e. according to a classification scheme, while the subject headings under these divisions are listed alphabetically. In the United States, the dictionary catalogue, which mechanically combines an author catalogue and an alphabetical subject catalogue, has taken root.

Bibliography is a special kind of scientific literature. It occupies a position of importance for science, culture and national economy being a means of bringing together, recording and classifying the contents of published and unpublished materials. Bibliography has its own object, purpose, techniques and form. Written works, duplicated by some means and designed for an indefinitely broad sections of contemporary and future readers, form the object of bibliography. An immediate purpose of bibliography is to influence the dissemination of written and printed records. The tracing and recording of written works, their selection according to some criterion, their description, sometimes supplemented with a critical appraisal, and their classification, constitute the techniques of bibliography. Listings and literature surveys are the forms of bibliography. This definition of bibliography - which, in our opinion, is the most felicitous of all existing ones - belongs to the distinguished Soviet bibliographer K.R. Simon (1887-1966).(1) This term, bibliography, as you learned from the first lecture,

---

1. Simon, K.R. - Bibliografija: osnovnye ponjatija i terminy. (Bibliography: basic concepts and terms). Moscow, Kniga Publishers, 1968, p. 26-35.

applies also to the discipline dealing with the elaboration of theory, methodology, and history of bibliography as a speical kind of scientific literature.

Bibliography, considered here as the last - but not least - type of secondary documents, holds a prominent position among other secondary documents by its importance, diversity and complexity. To describe the whole system of bibliography, even in brief outline, several special courses of lectures might not be enough, to say nothing of one lecture. That is why the only goal we can set ourselves here is to attempt a general outline of the most important types of bibliographies with a few illustrative examples.

According to their content and kinds of indexed materials, bibliographies can be general or special. General bibliography lists materials irrespective of their content and kind, while special bibliography covers only publications in a particular subject area, topic, or of a particular kind.

According to purpose, bibliographies can be enumerative, scientific, and recommendatory. Enumerative, or recording, bibliography registers, as exhaustively as possible, materials in different parts of knowledge, subject fields, or of specific character. Scientific bibliography aims to inform the specialists of the publications in their field of knowledge, theme, subject, or problem under study. Recommendatory bibliography has the objective of active reading guidance and of helping certain reader categories to choose literature that will suit them.

According to the time of publication, bibliographies divide into current, retrospective, and prospective. Current bibliography records materials as they are brought out, retrospective bibliography covers a certain time period prior to the time of compilation of the bibliography, and prospective bibliography reflects publications that are in press.

According to the place of publication of the literature indexed the following kinds of bibliography are distinguished: international, national and regional. International bibliography lists materials in different languages published in different countries. National bibliography covers materials appearing in a certain country or in a particular language. State bibliography is a variation of national bibliography; it includes materials published in one country. It is brought out by special State agencies, e.g. by book chambers in the USSR. Regional bibliography records local press, i.e. publications issued within a region, district, or any other locality.

Bibliographers distinguish many more types of bibliography according to different characteristics of division. We shall mention only two. Bibliography reflecting primary documents is commonly considered to be primary bibliography and is simply called "bibliography". Bibliographies listing bibliographies are considered to be secondary bibliographies and called "bibliographies of bibliographies", or guides to bibliographies. Work has long been in progress on compiling a bibliography of such guides, which would be a tertiary bibliography. Finally, according to the method of arrangement of entries bibliographies, as well as catalogues, may be classified, subject or author.

The types enumerated above by no means exhaust the existing types of bibliography. They were meant to give some general idea of the multiplicity of bibliographical sources and to help in the appraisal of specific bibliographical indexes used. There exist an enormous number of bibliographical indexes and aids. In the Soviet Union alone, thousands of titles of bibliographies and more than 150 special bibliographical periodicals are brought out every year. All bibliographies published in the world run well into tens or hundreds of thousands. Mastering all this wealth of information sources offers great opportunities to an incipient scholar. It is no chance that early European writers used a golden key as the symbol of bibliography and its motto read: "Omnium scientiarum clavis" ("Key to all sciences").

The most typical literature requirements of a specialist are the need to follow all current publications in his subject field and in related fields, and the need, which generally arises at the start and at completion of a study, to get familiar with all published sources pertaining to the theme of the study. The first need is met primarily by general current bibliographies, and the second, by special retrospective bibliographies. Since any particular bibliographic publication can be classified according to any of the characteristics listed above, we shall describe, by way of illustration, a few typical publications.

General bibliographies are typified by current enumerative national bibliographies. The Soviet current State bibliography is published by the All-Union Book Chamber as "annals". The weekly "Book Annals" provide details of current books and pamphlets, as well as standards, technical catalogues, abstracts of theses and other separately issued materials. All data appearing during the year in the "Book Annals" are cumulated with some abridgments in the "Bibliographical Yearbook of the USSR", printed in two half-volumes. Entries in both publications are filed systematically. They are provided with a subject index, an index of names, and an index of places. Details of newspapers, journals and magazines, bulletins, and serials are listed in the "Annals of Periodicals", which appear every five years with yearly supplements. Details of journal and newspaper articles are given by the weekly "Annals of Journal Articles" and "Annals of Newspaper Articles"; details of reviews and critical surveys printed in journals, newspapers and collections, are given in the quarterly "Annals of Reviews".

Similar publications exist in many countries. The popular American weekly "Cumulative Book Index", appearing since 1898, covers the books published in English in most of the countries of the world. Its specific feature is cumulativity, which means the inclusion of material from previous issues in subsequent ones. Cumulative volumes appear as substitutes for the April, July, September and December issues of the index and cover data for a period of, respectively, three, six, nine and twelve months and several years. The material is arranged as in a dictionary, with each book entered under its author, title and subject headings. The main medium of recording current British publications is the weekly "British National Bibliography", appearing since 1950. Details of new publications are arranged systematically; they are cumulated in quarterly issues and annual volumes. France boasts the oldest national bibliography - the weekly "Bibliographie de la France", published since 1811. It covers books, first issues of periodicals, theses, official and other publications. The material is arranged systematically; an author index and an index of titles of anonymous works are provided, as well as cumulative volumes printed every three, six, nine and twelve months. A special feature of this bibliography is that it announces books in press, which are subsequently included in the cumulative listings "Books for the Week", "Books for the Month", "Books for the Quarter", "Books for the Half-Year", Books for the Year". These lists are provided with indexes of authors and titles; the last one has, in addition, a subject index. The German Democratic Republic brings out the "Deutsche Nationalbibliographie", founded as early as 1931, which covers all publications appearing in the GDR, FRG and West Berlin, as well as books in German published all over the world. The "Deutsche Bibliographie", published in the Federal Republic of Germany since 1947, has approximately the same scope. Both bibliographies appear in several series as classified weekly issues which are cumulated every six months and five years.

Apart from these publications, designed for bibliographers, a range of general current bibliographical aids keep the scientists abrest of current literature. Details of the most significant books in all parts of knowledge which are published in Britain are provided by "Reference Catalogue of Current Literature" (Vol. 1-2, London, Whitaker, 1965). American books are listed by the annual "Books in print" (Vol. 1-2, New York, Bowker).

Details of periodicals can be found in some catalogues and lists:

Ulrich's Periodicals Directory: a classified guide to a select list of current periodicals, foreign and domestic. 10th ed. Vol. 1-2. New York, Bowker, 1963-1965.

Gregory W. Union list of Serials in Libraries of the U.S. and Canada. 3rd ed. Vol. 1-5. New York, Wilson, 1966.

British Union Catalogue of Periodicals, incorporating World List of Scientific Periodicals: new periodical titles. London, Butterworths, 1960 - (Monthly with quarterly, annual and 5-year cumulations).

Special bibliographies can best be described using as illustrations international and national special retrospective bibliographies:

Maizel R.E. and Siegel F. The Periodical Literature of Physics: a handbook for graduate students. New York, 1961. (American Institute of Physics).

World Nuclear Directory: an international reference book. 2nd ed. London, Harrap, 1963. 626 p.

Burman C.R. How to Find out in Chemistry: a guide to sources of information. Oxford, Pergamon, 1965, 220 p.

Bottle R.T. (ed.) Use of the Chemical Literature. London, Butterworths, 1962. 231 p.

Mason B. (comp.) The Literature of Geology. New York, American Museum of Natural History, 1953. 155 p.

Kerker A.E. and Schlundt E.M., Literature Sources in the Biological Sciences. Lafayette, Purdue University Library, 1961. 133 p.

Smith R., Guide to the Literature of the Zoological Sciences. 6th ed. Minneapolis, Burgess, 1962. 232 p.

Moore C.K. and Spencer K.J., Electronics: a bibliographical guide. London, Macdonald, 1961, 411 p.

Fry B.M. and Mohrhardt F.E. (eds.), A Guide to Information Sources in Space Science and Technology. New York, Interscience, 1963, 579 p.

Examples can be cited ad infinitum, but even the few presented above can give some idea of the nature of special retrospective bibliographies. The problem is then how to be made aware of the existence of such bibliographies on a discipline, theme or problem you are interested in.

Secondary bibliographies (bibliographies of bibliographies, or guides to bibliographies) serve exactly this purpose. They may supply information on the whole universe of knowledge or one part of it. Examples of such a guide are:

Besterman T.A., World Bibliography of Bibliographies. 4th ed. Vol. 1-5. Lausanne, Societas bibliographica, 1965-1966.

Malcles L.-N., Manuel de bibliographie. Paris, Presses universitaires de France, 1963. 328 p.

Totok W., Weitzel R., Weimann K.-H., Handbuch der bibliographischen Nachschlagewerke. 3. Aufl. Frankfurt am/Main, Klostermann, 1966. 3625.

Mention should also be made of Soviet bibliographies of bibliographies:

Krichevsky G.G., Obshchie bibliografii zarubezhnykh stran (General Bibliographies of Foreign Countries). Moscow, Publishing Office of All Union Book Chamber, 1962. 290 p.

Simon K.R., Istorya inostrannoi bibliografii (History of Foreign Bibliography). Moscow, Publishing Office of All Union Book Chamber, 1963. 736 p.

Gudovshchikova I.V., Obshchaia mezhdunarodnaia bibliografija bibliografii (General International Bibliography of Bibliographies). Leningrad, Krupskaya Institut of Culture, 1969. 105 p.

Novel indexes

The bibliographical indexes which have been considered above originated at different times and are long since in bibliographical practice. As is natural, they are used also in information work, although some of their shortcomings have come to be felt particularly acutely only recently. One of the main faults with conventional indexes is the relatively long time they take to prepare and produce. This is due to the fact that they are prepared exclusively by means of manual operations and do not lend themselves easily to mechanization and automation. Meanwhile, researchers and practical specialists feel a strong need for a bibliography that would have a short time lag between a publication and the appearance of systematic data on it, and would thus minimize one of the basic shortcomings of the abstract journal, its showness. Considering all that we have learned about the aging of scientific documents, we cannot but admit the fairness of this requirement. Such a kind of bibliography has come to be increasingly called current awareness service, because its purpose is to make scientists aware of all current publications in their respective subject fields.

It was quite clear that the necessary promptness here could be achieved only if modern technology, in particular electronic digital computers, was to be put to use. But, the biggest share of time and effort is spent in systematization, which as yet defies all efforts at automation. A way out was found in the development of new kinds of indexes in which systematization is performed automatically, without human intervention. The makeup and production of the new indexes will be made clearer if we go back to the methods of arrangement of index entries. We have seen that most indexes use classified arrangement, and some, arrangement by subject. Filing by authors is employed in a few bibliographies only, which are in fact printed catalogues reflecting the contents of several library stocks, i.e. union catalogues. Like any author catalogue of a library, they indicate whether a given publication is available, and in what libraries. Now it was discovered that alphabetization is rich in possibilities that had never been utilized in library catalogues and conventional bibliographical indexes.

Each word in the title of a document, composed by its author, is a potential subject heading, and the name of the author himself is a key to those bibliographical groupings which were treated in the previous lecture and which are established by means of citations. At the same time, punched card equipment and electronic computers have tremendous capacities for sorting and filing of any information collections by formal criteria, in particular for alphabetization. This fact was utilized in new kinds of indexes among which the widest use in information work is enjoyed by permuted-title and citation indexes.

Permuted-title indexes are based, by definition, on document titles. Experience shows that titles of the majority of documents in the fields of natural, exact and applied sciences are fairly accurate identifiers of document contents. In the preparation of a permuted-title index, a computer program, using a so-called stop-list, identifies in titles and skips all common words (syntactic words, such as pronouns, numerals, conjunctions, prepositions, particles, auxiliary verbs) as well as general nouns and adjectives (e.g., "problem", "study", "method"). The rest of the words and terms appearing in document titles are used as subject headings. They are called keywords and are displayed in the index in the context of their title. That is why such indexes have been termed KWIC (Key-Word-In-Title) indexes.

All keywords contained in abstract titles are alphabetized vertically in the search column which is justified on the left, with the rest of the title words displayed to the left and to the right of the keywords. The index-line generally does not exceed 80 characters in length, including spacing. Therefore, if the text following a keyword does not fit into the right-hand part of the index-line, while there is room in the left-hand part, then the line is "wrapped around", i.e. transferred to the left.

If the title is still too long, it is truncated to match the length of the index-line. The end of the title is marked by an asterisk, which helps the title to be read correctly. If the title ending is cut off, the whole of the title can be reproduced in other parts of the index, which are easily located through title keywords. A title usually comprises not less than 4 or 5 keywords and therefore is displayed in the index 4 or 5 times minimum. Because title words appear in the index in different groupings the index is called a permuted index, although it should more aptly be called a rotated one. It is to be remembered in using a permuted-title index that the same concepts can be physically separated in it if they are expressed by synonymous words. For example, if you are looking for materials on transistor triodes, in addition to titles containing the word "transistor", you will have to scan all titles containing the synonymous term "semiconductor" and its derivatives. The strategy of searching in a permuted-title index is easy to learn and usually presents no difficulty in practice.

Permuted-title indexes may be used as current-awareness publications or as auxiliary indexes to abstract journals and other secondary publications. In the former case they are provided with a name and a bibliographical index, and the codes located in the right-hand column denote the title abbreviation, year, number and the first page of the periodical in which the respective article appears. In the latter case these codes stand for the ID numbers of the abstracts in the abstract journal. In either case, these codes will be used to locate the entry which provides full bibliographical description of the publication.

Permuted-title indexes were first produced in the United States in 1958. They owe their introduction in information work to the late H.P. Luhn, the American engineer who is credited with being the principal originator of the method. The Chemical Abstracts Service publishes since 1961 the semi-monthly current-awareness index "Chemical Titles", which announces new articles from 600 major journals in chemistry and chemical technology. Each issue comprising details of 3,000 papers and including an author and bibliographical indexes, takes 5 hours of computer time to compile. Another instance of an extensively used permuted-title index is the index to the abstract journal Biological Abstracts. Called B.A.S.I.C. (Biological Abstracts Subjects in Context), it appears semi-monthly since 1962. It replaces the former subject index and is appended to every issue of the abstract journal. Its yearly coverage is about 120,000 abstracts. Regularly produced permuted-title indexes are scheduled for the nearest future for individual series of the VINITI Abstract Journal.

Citation indexes constitute a basically new kind of index; they supply answers to such requests of researchers and specialists as cannot be met by conventional indexes. The main question which they answer is in what new publications the former works of particular authors are cited. By cited works are meant those works to which bibliographical references are made in the texts of other publications. These indexes embody the bibliographic coupling method which was discussed in the last lecture.

The first, largest and best known index of this kind is the Science Citation Index (SCI) regularly produced since 1964 by the Institute for Scientific Information (U.S.A.). In 1966, it covered 12,500 issues of 1,573 journal titles in all exact, natural and applied sciences. It included 3,000,000 references to former publications of different years contained in the 300,000 publications which appeared in 1966. The index is composed of two parts: the index proper (Citation Index) and the Source Index. The indexes appear in four issues each - three quarterly issues and one annual issue. The cumulative volume of the Citation Index for 1966 comprised 5 volumes, and the Source Index, three.

The material in the Citation Index is arranged alphabetically under the names of the authors of cited documents; after each of these the authors of the source (i.e. citing) documents are listed, also alphabetically. Details on works of the same author are filed chronologically. Following the name of the author, the source of a publication is given, including an abbreviated journal title, year of publication of the cited or source document, volume, and the number of the first page. Preceding the year of publication, a code is printed describing the source document type, viz., abstract, review, bibliography, patent specification.

In the Source Index, the material is also arranged alphabetically under first authors. In contrast to the Citation Index, where the first author alone is entered, here all the joint authors are given, if they are not more than ten. A specific feature of this index is the availability of the full title of an article or patent specification. It also gives the abbreviated title of the journal in which the article is published, the volume and page number, year of publication, the code for the source document, the number of citations, the issue number, and the ISI accession number.

Document searching in this index is conducted in the following manner. A starting (or "target") reference is first established, which is known to the searcher beforehand, or obtained from conventional sources (encyclopaedias, textbooks, handbooks, guides, bibliographical indexes, abstract journals, library catalogues.) Under its author, details of all the documents related to it are found in the Citation Index. Full entries for the source documents are found in the Source Index, and their relevance to the query is assessed. If these documents are not sufficient, any of the earliest documents citing the target reference is used as a new target reference for further searching. This cycle is repeated until all the necessary documents are identified or until the searching yields no more information.

The technique of searching in a citation index can be illustrated using as example a search for documents on nucleic acids, which was instanced in the previous lecture (Fig.4). The 1941 paper by E. Rabinovitch and L. Epstein which was taken as the target reference, cites five other works. If the earliest of the references listed under the first work (L. Mikhaelis 1947 and 1950, filed one after the other) are taken as target references for further search, they will provide details of four more works on this subject. Finally, using V. Tsanker's work as the target reference will help to identify three more papers. Thus, by knowing only the name of the originator of this theory, 13 out of 15 works on the subject published between 1941 and 1960 can be found in three searches.

The most important and most typical trait of a citation index is that it provides answers of a thematic nature, although the entries are arranged alphabetically under authors. All conventional indexes designed for this purpose require content analysis of each indexed document. This is the reason why they cannot be automated as yet. The Science Citation Index is fully computerized, because formal rules are used for its compilation: automatic grouping of material by subject-matter is effected, as has been said before, by the principle of bibliographic coupling of documents.

The idea of such an index was first put forward by E. Garfield, now director of the Institute for Scientific Information, as early as in 1955. It was suggested to him by the Shephard's Citation (Chicago, 1873-) published in the United States since 1873, and in part also by bibliographies of reviews which use the same principle of material grouping and have been a standard feature of bibliography since the 18th Century. (The first bibliography of reviews appeared in Paris in 1771-1772). The importance of citation indexes is by no means confined to thematic searching. These indexes helped to discover many scientific publication patterns, which were discussed in the last lecture. They allow to trace the evolution of a scientific idea as it is reflected in the publications of different authors, help to establish how a problem stated in some work has been treated, whether an author's theory has been supported, whether it found application, and whether it was original.

All this assumes special importance in the context of the interpenetration of sciences, which is a sign of the times. If a work of an author in a given field of knowledge is cited by many workers in other fields, it points with certainty to the fact that the ideas expressed in this work have proved fruitful for a number of scientific disciplines. The important point is that the establishment of new links between disciplines and theories is detected and recorded by citation indexes with a promptness which cannot be achieved by any other means, and irrespective of the quality of any existing classification schemes or the skills and subjective judgments of indexers. In a citation index, the relationships of a given idea expressed in a document are established by the scientists themselves, viz. authors of the publications citing this document. The number of references to one or another document provides an objective basis for the assessment of the significance and viability of the ideas it contains.

Current-awareness information

A form of rapid information, called current-awareness service, has been gaining popularity recently. It consists in timely (a maximum of one month following publication) alerting of information users to the appearance of useful materials and is accompanied by a duplicating service supplying prompt and not too costly copies of any of these materials.

According to the organization of the searching apparatus, two types of current-awareness services can be distinguished: information and bibliographic publications, and selective dissemination of information. Information and bibliographic publications present data in a definite order and in various aspects - by subject, by author, by language, by time of publication, etc. - using the appropriate indexes.

The most extensively used current-awareness publications are classified or alphabetical subject indexes, and permuted-title indexes, discussed above. A specific form of current-awareness service are contents-lists, compiled from the contents of the recent issues of the major journals in particular subject areas. The most popular of these publications, "Current Contents" (USA) includes an index of authors and their addresses and appears weekly in three series covering in excess of 2,000 journal titles in the natural sciences.

With selective dissemination of information, an information user entrusts to an information service the searching for the information that might interest him. He specifies his profile of interests which may include class marks, subject headings, descriptors, names of authors of the works sought or their references. All the materials accessioned by an information service are checked for these characteristics, and, in case of matching the user interest profile, he is supplied with the specific data that interest him. The feedback to the user provides for constant adjustment of document selection, thereby improving the personalized service.

A typical example of selective dissemination of information is the ISI's weekly service Automatic Subject Citation Alert (ASCA). This service notifies individual subscribers of the details of all the publications contained in the 1,500 journals received by the Institute which match the title keywords or names of the authors of cited works, i.e. the characteristics included in their profiles.

We have considered the most important kinds of secondary documents, information and bibliographic publications. They remain the main sources of data on scientific literature. This lecture is only the first step in getting familiar with them. It is our wish that at all stages of your scientific work these tools should receive your most considerate attention. Familiarity with these sources and their skillful use are major elements of a scientist's qualification. But treat them critically, for along with science as a whole, they too are undergoing revolutionary changes. Novel index forms bear witness to it. The next lecture will deal with the activities of information agencies and libraries, which are responsible for the preparation and publication of secondary documents. Further lectures will detail the principles and procedures of their preparation and use.

Checkup Questions

1. What are the typical traits of a modern encyclopaedia?
2. What is the role of surveys of science and technology?
3. What is the structure of an abstract journal and what are its functions?
4. What principal answers are provided by catalogues of libraries?
5. What are the existing kinds of bibliography and what purpose do they serve?
6. What is a permuted-title index and how is it used?
7. What new opportunities are offered by citation indexes as compared with conventional indexes?

Literature

1. Carey, R.J.P. Finding and Using Technical Information. A handbook for students, teachers, and practising engineers, London, Arnold, 1966. VI, 153 p.
2. Fowler, M.J. Guides to Scientific Periodicals. Annotated bibliography. London, Library Association, 1966. XVI, 318 p.
3. Davidson, D.E. Periodicals: a manual of practice for librarians. 2nd ed. London, Grafton, 1964. 200 p.
4. Luhn, H.P. Keyword-In-Context Index for Technical Literature (KWIC Index). - "American Documentation", 1960, vol. 11, N.4, p. 288-295.
5. Garfield, E. "Science Citation Index" -- A New Dimension in Indexing. -- "Science", 1964, vol. 144, p. 649-654.

4. INFORMATION AGENCIES AND SPECIAL LIBRARIES - AUXILIARY INSTITUTIONS OF SCIENCE

Any study or R & D work requires certain knowledge on the part of the specialist concerned. A portion of this knowledge has been acquired by him during training, the rest accumulating in the course of his research and practical work, through the continuous contact with life and production, as well as, in the course of his life-long self-education and raising of his qualifications. This knowledge determines the skills and constitutes the experience of a scientist or practical worker. But however high the qualifications and broad the experience, they are not sufficient in themselves. To solve any scientific problem, the latest data on what has been and is being done on it are an absolute necessity. In other words, initial scientific information has to be obtained.

Channels of dissemination of scientific information

We have familiarized ourselves with the manner in which scientific data are spread in a modern society, and in the two previous lectures have considered the system of scientific publication as the main medium for the communication of scientific knowledge. Today we shall discuss this problem in a somewhat different aspect and find out how scientific information is channeled, i.e. what routes it takes from the scientists who generate it to those who use it. Particular attention will be paid to the activities of those ancillary institutions of science which are positioned along these routes.

Personal contacts and correspondence had long been the principal information channel. It is for this reason that the conventional form of a letter from one scientist to another has for many years endured in scientific publications. Personal contacts between scientists remain to this day one of the ways of information transfer. They are supplemented by the exchange of offprints and preprints of journal papers, by participation in scientific congresses, conferences and symposia. Recent investigations in the science of science, in particular those using the method of bibliographic coupling, testify to the importance of these channels of information transfer. It has been found that the front of scientific research divides into small stretches, each corresponding to the work on a particular problem conducted in parallel by a small (100-150 persons) team of leading researchers in different countries and establishments, which form a sort of corporation, or "invisible college", as it was termed. Personal contacts, direct exchanges of information are of great significance for the members of these invisible colleges.

However, at the present time the bulk of scientific information is disseminated through scientific documents.

Fig. 14 is a block diagram of the main channels of distribution of scientific information from generators to users. The outward lines represent personal contacts among scientists, their correspondence, exchange of offprints and preprints, as well as participation in congresses, conferences and symposia. The inner lines represent the distribution of unpublished documents, which after having received the approval in scientific or technological institutions where their originators work are normally made known to the users by the scientific information agencies. Dissertations, defended in research and academic institutions, are made available to users through special libraries.

The bulk of scientific documents, it will be recalled, consists of the published records. The channels of diffusion for these are at the centre of the diagram. The generators of these records, namely, the authors of books, periodical articles or other materials ready for publication, pass them over to the book publishing houses or editorial offices of journals, as the approving bodies. Frequently, an author hands his manuscript to the publisher after a discussion and on recommendation from a scientific or educational institution where he is employed, but this is not reflected in our diagram.

Publishing house and printing office operations

In order to understand how scientific documents are published, it is necessary to get a notion, if only superficial, of the major steps in the publishing and printing routines. The manuscript of any work handed over by its author to the publishing office is reviewed to evaluate its theoretical and practical worth. The manuscripts accepted for publication are thoroughly edited from the scientific and literary points of view. The editor sees to it that the author's ideas are stated clearly, the terminology and formulations are correct, and the style is concise; he verifies as far as possible the reliability of the facts and figures cited. After the editing the manuscript is retyped. It is then read out by the proof reader, who is responsible for the spelling and punctuation, as well as by the copy preparer who decides all typographic features of the future publication: the type size and face, the method of illustration reproduction, the set of type and paper size, the quality of the cover and binding. This completed the editing routines, and the manuscript proceeds to the printing office.

The complex typographic process can be broken down into three steps: preparation of the form, or composition; printing; and binding. Composition may be performed by hand, but more often it is done mechanically, on line-setting machines, or linotypes, and monotype composing machines, or monotypes. The composed lines are put together in galleys, from which a trial print is made, read by the printer's readers or sent to the publisher for reading. After making corrections in the galleys, the makeup follows, i.e. the type pages of the same size are formed from the setup type. The proof sheets of the makeup are read again in the publishing office (and usually also by the author), and the composed matter is corrected according to them. This procedure is repeated for the third time as the so-called revision, following which the form is considered to be ready for the press. The form is used in the printing, i.e. in the preparation of the necessary number of prints. The work on a publication in the printing office completes with the binding i.e. finishing of the printed pages. They are folded into signatures, usually 16 pages large, collated into an unbacked book, and stitched. The unbacked book is trimmed on three sides, provided with a cover or inserted into the binding.

After a book has been produced, its run is forwarded to the bookshops, while journals are mailed to the subscribers. It is self-evident, however, that the buying of books for personal libraries and the individual subscription to journals can only satisfy a small fragment of the scientists' and practicians' information needs. To put the case, the limited resources of a researcher have only to be compared with the colossal and ever growing volume of scientific literature.

Relationships of the information agencies and the libraries

This explains why the information agencies and special libraries serve as the major auxiliary institutions of science, discharging the functions of intermediaries between the generators and the users of scientific information, and providing for the normal operation of information channels through which scientific documents are spread.

The tasks of these institutions have much in common, as they correspond in the main to the tasks and stages of scientific information activities which were formulated in the first lecture. Their mission is to identify and collect scientific information, to analyse and process it into forms suitable for storage and searching, to provide for its prolonged storage and dissemination, including retrieval on demand. But each of these institutions performing information functions has certain peculiarities, and different particular tasks solved by different subject specialists employing different techniques. Today we will reveal these differences in order to learn how to better use the opportunities offered by the information agencies and libraries.

One of such differences related to the staff of these institutions, was succinctly stated by a group of leading American scientists in a report to the President of the United States: "In short, knowledgeable scientific interpreters who can collect relevant data, review a field, and distill information in a manner that goes to the heart of a technical situation are more help to the over-burdened specialist than is a mere pile of relevant documents. Such knowledgeable scientific middlemen who themselves contribute to science are the backbone of the information center; they make an information center a technical institute rather than a technical library".(1) Thus, an information specialist is a specialist in a certain field, a researcher participating in the solution of a scientific problem, who provides the research team with information. On the other hand, a librarian -- who may also have a training or even qualifications in one of the broad fields of knowledge -- is primarily an expert in reader psychology, proficient in teaching methods, and capable of seeing his way in the multifarious book production environment. An information specialist and a librarian are experts in different fields.

The ultimate goal of an information service is to notify of new facts and ideas, to provide answers to questions about facts, e.g. What are the properties

---

1. Science, Government, and Information. A report of the President's Science Advisory Committee. Washington, 1963, p.33.

of the given object? What objects possess the given properties? What is known on this problem? Since answers to these questions today are provided by scientific documents - books, journals, patent specifications, technical reports etc., - information agencies handle these and will frequently supply only references, not data, e.g. tell what facts are to be found in certain documents or what documents contain the wanted information. Hence, working with documents is not the primary task of the information agencies, but only a means of manipulation of scientific information contained in them, in order to obtain new data by its logical processing. The library is an educational and ancillary scientific institution providing for public use of the library materials and designed to help the reader in choosing his books, and to guide him in his reading. Special libraries are an important link in the information service of science, and one of the most important auxiliary means of elaboration of strategical problems of science, national economy and culture. Therefore, these auxiliary institutions of science have differing ultimate objectives, although now they work with largely the same scientific documents.

Another difference pertains to the individuals to whom the work of these institutions is addressed. While these are the selfsame persons, namely, scientists and practical workers, they are called differently by the information scientists, who use the term "information users", and librarians, who use the term "readers". It should be admitted that the existence of different terms is justified to the extent that they express different notions. An information user refers to an information centre or its publications in order to obtain specific data which are necessary for his research work. He has a right to demand an accurate and complete answer to a specific query from the information specialists that keep track of the advances in this specific field by means of the documentary sources. In the library, the same person appears in the capacity of a reader, who is to be provided, besides the materials pertaining to the topic of his immediate concern, also with books which would broaden his world outlook, improve his skills, acquaint him with related parts of knowledge, and satisfy his general scientific and broad social interests.

And finally, the institutions under discussion operate in different modes. While this difference is not one of principle, it is nevertheless apparent in practice, and is probably not accidental. Most information agencies provide their services on a regular basis, which corresponds to the rate of advancement of a science, to the frequency of appearance of new publications and unpublished materials. Information agencies attempt to cover as fully as possible the new documents, to effect their analytico-synthetic processing, and then to bring their contents by equal portions to the scientists in the shape of separate issues of periodically appearing abstract journals, surveys, express-information bulletins or bibliographies. True, parallel with this the information agencies will preserve the documents for prolonged storage and searching and retrieval on the users' requests. But this work occupies a secondary place in the majority of information agencies. The libraries, on the other hand, operate in the "question - answer" mode. They accumulate materials, reveal their contents, and organize them for storage in such a way as to be able to produce them on request. It is a known fact that almost every library regularly informs its readers of the accessions, but it is also known that this work is by far more limited in scope than their on-demand service to readers.

Organization of scientific information activities in the USSR

Information work in the Soviet Union is based on the principle of centralized processing of the world's scientific and technical information and its dissemination to the different branches of national economy. This means that all kinds of scientific documents are processed in centralized State agencies, with their output in the form of abstract journals, express-information bulletins, surveys, bibliographical indexes and printed cards distributed among the various research institutes, design offices, industrial enterprises, institutions of higher learning, as well as branch and regional informational agencies. On requests from these institutions and offices, the centralized information agencies furnish copies of the original primary documents announced in the secondary documents that have been enumerated above. This flow of scientific and technical information is usually called a downward one. Another downward flow of information to the information centres of enterprises and institutions is generated by the branch-oriented information bodies in the form of newsletters or information leaflets, surveys and other documents in the respective fields of science, engineering and national economy, as well as by the territorial (republican and regional) bodies which put out state-of-the-art reports on problems of topical interest to the enterprises and institutions of the given region, and channel scientific documents in local (other than Russian) languages.

This flow is countered by the information blowing in the opposite direction, which was termed the upward flow of information. It consists of R & D reports and information cards with details of the studies conducted, which are directed by research institutions and educational establishments, design offices and industrial enterprises to the All-Union Scientific and Technical Information Centre and to the branch-oriented information centres for registration and storage.

All information agencies in the Soviet Union thus fall into four large groups: all-union (national), branch-oriented, regional, and in-house centres serving the needs of enterprises and institutions. The all-union information agencies include the All-Union Institute of Scientific and Technical Information (VINITI), the All-Union Scientific and Technical Information Centre, (VNTIC), the Central Research Institute for Patent Information (CNIIPI), and the All-Union Research Institute for Technical Information, Classification and Coding (VNIIKI). The branch-oriented information agencies include a few dozen central institutes for scientific and technical information (CINTI) and technical and economic research in the respective areas of science, technology and national economy, each of which has its own central scientific and technical library (CNTB). The regional information agencies include 15 republican institutes of scientific and technical information and popularization with their respective republican scientific and technical libraries (RNTB), as well as several dozen central bureaux of technical information (CBTI) in the regions, with their central scientific and technical libraries (CNTB) and houses of scientific and technical popularization (DNTP). Finally, the information bodies at enterprises and institutions comprise scientific and technical information departments (ONTI) of scientific research institutes (NII), design offices (KB) with their own scientific and technical libraries (NTB), as well as technical information bureaux (BTI) with technical libraries at enterprises. There are about 5,000 of such

ONTIs and BTIs. The same functions are performed by tens of thousands of separate technical libraries functioning at enterprises, and academic libraries at higher educational establishments.

The general supervision over the scientific information activities is exercised by the State Committee for Science and Technology of the Council of Ministers of the USSR. The ministries and departments organize and control the work of the corresponding branch-oriented information agencies.

The All-Union Institute of Scientific and Technical Information (VINITI) of the USSR Council of Ministers State Committee for Science and Technology and the Academy of Sciences of the USSR was organized, as has been mentioned, in 1952. In the 17 years past, it has grown into the world's largest information centre. It employs some 2,500 staff workers and more than 22,000 supernumerary research workers. The volume of its publications exceeds 30,000 publisher's signatures a year, which ranks it with the largest publishing houses of the Soviet Union. The VINITI has great typographic resources which include its Publishing and Printing Combine with a staff of over 1,700. The main functions of the Institute are as follows: the study and processing of the world's scientific and technical literature in all parts of knowledge; the preparation and publication of abstract, survey and bibliographical information; the duplication of the original publications on requests from institutions, enterprises and individuals; the organization of information and reference services; the co-ordination of translation activities in the country; the publication of collections of digests of articles of topical interest ("Express Information"); the conducting and co-ordination of research in informatics; the organization of advanced training courses for specialists employed in information bodies.

The VINITI can serve as a good example to show the internal workings of a large information centre, and the procedures of preparation of abstract publications. The daily batches of material coming to the Institute are forwarded to the department of preliminary scientific processing, where the scientific worth of each publication is assessed, and the relevant series of the Abstract Journal are indicated where it is to be indexed. The publications selected are marked with special symbols, and passed over to the bibliographers who prepare them for description by underlining the necessary bibliographic data. A card is prepared for each publication which accompanies the publication throughout its processing cycle. Card copies are added to the catalogues and auxiliary indexes. Because most of the publications go simultaneously to several Abstract Journal series, the corresponding number of their copies is made, while the original is stored in the department of scientific resources.

After the department of preliminary scientific processing and the bibliographic group, publication copies with cards are distributed among the subject departments where they are assigned to the staff and extranumerary abstractors, according to their type, content, and language. The abstractors return the copies of the publications together with an abstract, which is then thoroughly edited and checked for errors in the corresponding subject department. Then the abstracts, classified according to a special list of subject headings, are grouped for the next Abstract Journal issues; following a discussion at the editorial board, the manuscript of the journal is sent for publication to the Publishing and Printing Combine. There the usual publishing routines already discussed in the beginning of the lecture are performed. You may imagine what high level of discipline and co-ordination must be maintained by the numerous persons participating in these processes in order to complete them in four months' time, which is the optimal information delay, as you will recall. And these are the techniques and procedures of only one of the various functions of the VINITI.

Now let us familiarize ourselves with the functions of the other national information agencies. The All-Union Scientific and Technical Information Centre (VNTIC) of the State Committee for Science and Technology of the Council of Ministers of the USSR at the present--initial--stage of its development is charged with the following: the State recording and bibliographical control of the R & D projects carried out in the country; the maintenance of a file of R & D documentation; the publication of regular information on technical reports accessions; the reference and information services.

The Central Research Institute for Patent Information (CNIIPI) of the USSR state Committee for Inventions and Discoveries is charged with the following: the scientific processing of the world's patent materials; the preparation and publication of current-awareness and survey information for the field (abstracting services are provided by the VINITI); the information provision for patent examination work; the reference and information services on the domestic and foreign patent materials; the publication of information materials on the problems of invention and patent work. The All-Union Research Institute for Technical Information, Classification and Coding (VINIIKI) of the USSR State Committee for Standards, Measures and Instrumentation performs the following tasks: the scientific processing, preparation and publication of analytical information on standards; the reference and information work on standards and interdepartmental norms; the conducting and co-ordination of research projects on classification and coding systems; the organization of a State Service of Standard Reference Data (GSSSD); the publication of information materials on standardization problems.

The branch-oriented information agencies are represented in the Soviet Union by the central institutes of scientific and technical information (CINTI) and technical and economic research under the different ministries and departments, as well as by the scientific and technical information department (ONTI) of the leading scientific research institutes (NII). They have the following objectives: to ensure the scientific processing of information on the advances made in the appropriate organizations, institutions and enterprises in the corresponding fields, to prepare on its basis current-awareness, abstract and survey information; to provide reference and library services over the range of specific subjects of the field; to conduct analytical survey and comparative technical and economic studies into the problems of the development of the field, and some others. They lay particular emphasis on a systematic study of the technical and economic performance indices of industrial enterprises, of the design and operative technical and economic data on products and technologies, as well as of the methodology of organization of production, research and development work.

Regional information agencies, as has been said before, are represented by different types of institutions. Republican institutes of scientific and technical information and popularization provide the following services: the information services for the Party and government bodies of the Union Republics; the reference, information and library services to the enterprises and institutions within the respective republic; the dissemination of interdisciplinary information within the republic; the preparation and publication of information materials in the corresponding national languages; the co-ordination and conducting of scientific and technical popularization in the republic, and some others. The central bureaux of technical information (CBTI) and houses of scientific and technical popularization (DNTP) in regions, cities and other localities carry out similar functions within their respective administrative areas.

The information bodies of enterprises and institutions are represented by scientific and technical information departments (ONTI) or bureaux of technical information (BTI) et enterprises, research institutes (NII), and design offices (KB). They are entrusted with the following tasks; the prompt and pinpointed information and library services to the scientists, specialists and industrial workers; the popularization of progressive scientific and technical experience; the control over the utilization of information materials; the supply of scientific and technical information and documentation to the corresponding all-union and central branch-oriented information agencies. Such is, presented in very general terms, the framework of information services and the missions of the main information agencies in the Soviet Union.

Special (research) libraries

Now we must get a bird's eye view of the libraries, another important type of auxiliary institution of science. According to function, all libraries are grouped into public and special (research) libraries. Public libraries provide their readers with the most significant literature which is necessary for the all-round development of man, for his education, and mastering of general and polytechnical knowledge. They are not oriented towards any particular reader category and engage in active populatization of books and in reading guidance.

The mission of the special (research) libraries is to assist in research and development activities, in industrial and agricultural production, in the advancement of construction, transport, public health, communications, social and cultural life of the country, as well as in the training and upgrading of the skills of the personnel in these fields, by providing for public use the library materials, by guiding specialists in their reading with the aim of improving their ideological, political, scientific, professional and cultural levels. These libraries belong to the four principal types: (I) large independent scientific libraries of universal character; (2) libraries of research institutes; (3) college and university libraries, and (4) industry-oriented libraries.

Large independent universal libraries are a special type of library. Along with meeting the information needs of the readership at large - specialists in science, industry or culture - they also discharge cultural and educational functions which are typical of public libraries.

Libraries of research institutes provide scientific literature to their research teams.

College and university libraries, apart from serving scientists, assist in the educational process and play a significant role in the training of specialists. University libraries hold a special position in that in keeping with the specifics of university education, they help the students in acquiring a broad scientific background.

Industry-oriented libraries help practical specialists in resolving their production tasks. They form a majority of the special libraries and include technical, medical, agricultural libraries, as well as libraries of transport and communication institutions, construction and similar organizations. In those institutions and organizations which have scientific information centres of their own, libraries form subdivisions of these; where such information centres are lacking, their functions are performed by special libraries.

Library routines

With all the diversity of library types, their structure and operations are basically the same and can be described in a generalized form. The main library routines are the acquisition of library materials, their cataloguing (or scientific processing), the organization of library holdings, and service to readers. To these four functions correspond the principal divisions of a special library which are, as a rule, functionally structured: the acquisition department, the cataloguing (or processing) department, the stack department, and the service department. Many large libraries also have a bibliographic (or bibliographic reference) department, research and methodological department, administrative and other departments. As the Indian library scientist S. Ranganathan has aptly put it, "Library is a growing organism". Transfers of large masses of literature are taking place in it daily. Two flows are clearly discernible:one consists of the new books and periodicals received by the library, and the other of the literature issued on requests from the readers and returned after having been read. The first flow has been termed "the way of the new book", and the second, "the way of the reader's request".

The starting point in "the way of the book" is the acquisition department. There, on the basis of a special document stating in general terms the reader's demands, which is called "acquisition policy", it is decided which of the new publications should be acquired by the library. The librarians learn of the appearance of these publications from current bibliographies. The sources of acquisition are bookshops and specialized library institutions - book supply agencies - carrying out centralized acquisition, as well as other libraries with which book exchange is maintained. The purchased books are accessioned, after which they are considered as added to the stock. The accessioning consists in marking a new item with the library stamp and an accession number. The accessions records are used in the individual and summary stock-taking of accessions.

Now the accessions go to the cataloguing (processing) department. There a card is prepared for each item, containing its main bibliographical details, and its subject matter is briefly identified by means of conventional symbols -- class marks or subject headings. These symbols are inscribed on the book and the card, following which the book proceeds to the stacks, while the card is reproduced in the necessary number of copies. The cards are put together to form catalogues of libraries, described in the previous lecture. The system of catalogues in a library, which may comprise up to several dozen catalogues - author, classified and alphabetical subject - introduces both the readers and the staff to the library holdings.

The literature is filed in the stacks in a certain order. It varies according to the principle of arrangement of the book stock (or shelving arrangement) adopted in the library. In case of a classified arrangement, it resembles the filing of entries in a classified catalogue. The books are filed by parts of knowledge according to their subject matter. However, unlike a catalogue entry which can be placed under several divisions corresponding to the different subjects dealt with in the book, the book itself can only occupy a single place on the shelf. Classified arrangement is advantageous for a small library where the book stock does not exceed a few tens of thousands volumes. The librarian is enabled then to give advice to a reader straight from the shelf, he can more easily memorize the location of each book and the general layout of the stock. The library can provide open access to the shelves, that is let the readers make their own selection among the books on the shelves. True, even in this latter case, catalogues offer a more effective means of interpretation of the book collections and recommending books, since they reflect the books in several aspects, as well as those books which are off-the-shelf at the moment. Most public libraries use the classified shelf arrangement of books.

In larger libraries, whose holdings may amount to several hundred thousand volumes, the above-mentioned advantages of classified arrangement will fade, while its shortcomings will become more pronounced. Spare place for accessions has to be provided in each section, class and subclass, and when it is found to be still lacking, the labour-consuming shifting of the holdings has to be made. Because books of the same content vary widely in height, the vertical shelf spacing must be such as to accommodate the largest book. This results in stack space being lost both vertically and horizontally, as a librarian would say. For this reason the larger special libraries increasingly adopt one of the formal arrangements, which take into consideration certain formal characteristics, and not the content of the books. The most popular method is shelving by size, when books are ordered on the shelf according to their height (size) and the date of arrival in the library (accession number). Such an organization of the stock saves much space in the stacks and ensures a prompter and more accurate issuing of books to readers. In this case a book's call mark (or its location mark) will not be its classification number, as with classified arrangement, but its format mark and accession (or serial) number.

Filing of the cards for accessions in catalogues, and of the new items in the stacks completes "the way of the book". Now these publications are ready to be used. A reader, when he comes to the library, finds out what publications he wants. To do this he makes use of secondary documents: bibliographical indexes, card indexes, abstract journals, card catalogues. In the catalogues the reader finds out whether the wanted publications are in the library holdings and looks for their call marks. He is assisted in his searching by the bibliographic reference department of the library or by the bibliographer on duty. He then fills in his requests for the books, giving the main bibliographical details and the call marks. This starts the second library flow, termed "the path of the reader's request". The request form, sometimes checked for possible errors in a special division, is then sent to the stacks, where the librarians remove from the shelves the books sought. These are forwarded to the delivery station from which the reader takes them for his use. The time lag between the moment of handing in the request to the moment of book issue to the reader may vary in different libraries all the way from a few minutes to several hours.

Provision of library materials to the readers is effected in two ways. A reader can either use them in the reading room, or take home. In the former case, the library literature is issued for the period until the closing of the library and reserved for this particular reader on special shelves for several consecutive days. In the latter case, the reader is issued his literature for a fixed period of time ranging from one week to one month in the lending (circulation) department. In some libraries the books issued in the reading room, and those loaned to the readers have different physical locations in the stacks and are reflected in separate catalogues. As a rule, the delivery station or the circulation department will keep a file of reader's cards, in which the issued books are recorded, and a file of book cards in which the names or numbers of the readers who are issued a given book are marked. Thus the librarians at any one moment can tell what books are borrowed by a given reader, when he must return them, who is reading the book which is off-the-shelf and requested by another reader. On return of the loaned books, they are restored in their places in the stacks. This completes the second continuously circulating flow of library materials. It should be added here that if the books requested by a reader are not available in a library, they can be ordered in other libraries within the country or even abroad using the so-called interlibrary loan service, which links together many special libraries.

Thus, special libraries are important auxiliary instititulons of science charged with providing scientists and practical workers with the necessary scientific documents. Ideally, every such library should give prompt and complete answers to a request asking what documents exist on this or that subject or theme, and after the reader has specified the type of the wanted materials, time of their publication, their language and the aspect of treatment of the corresponding subject, should submit these documents for the reader's use. In actual fact, special libraries are far from such a happy state of affairs. Not a single of them, not even their totality in any one country, can hope to accumulate all the world's scientific literature, to say nothing of all the scientific documents. Moreover, although libraries attempt to cover all bibliographical publications, the sheer bulk and diversity of these materials will not permit a timely and complete coverage of all the data on the existing scientific literature in a given subject. Catalogues of libraries reveal only the main items of scientific document contents and are not meant for multi-aspect literature searching. The majority of basic library routines are performed by hand, they are poorly mechanized and totally lack automation. The time it takes to locate, retrieve and use library materials by very far surpasses the time that a researcher or practician is prepared to spend for his information work. Hence a certain dissatisfaction with the present-day special libraries and the persistent efforts to develop techniques and means of improving their operation, which are frequently labelled, not too felicitously either, as the techniques and means of creating "the library of the future". It would be more appropriate to speak of the necessity of creating a truly modern library.

A Universal Information System and the developing countries

It is a very important task to coordinate the information activities throughout the world, to organize a wide exchange of primary and secondary documents, programs for machine processing of texts in natural languages, uniform information media and so on. But this is a very difficult problem.

The organizational principles and the build-up of a national system of scientific and technical information are entirely determined by historical, economic, social and other conditions prevailing in the country concerned. Technical means employed in scientific information activities depend on the country's area and population, level of technological development, economic conditions and a great many other factors. For that reason any attempt to make general recommendations that would be valid for a majority of nations or universal are doomed to failure.

This is particularly true of developing countries. In this connection J. Dean from the Institute of Librarianship, University of Ibadan (Nigeria) said in his review of L. Asheim's book "Librarianship in the developing countries" (Urbana, University of Illinois Press, 1966. 95 p.) : "The main limitation of Dr. Asheim's book is implicit in its extremely broad approach. He deals neither with particular developing countries, nor with a group of developing countries, but generalizes about all developing countries as though they possess to a great extent a common identity. This is not so. These countries comprehend a vast area of the world; the whole Latin America and most of Africa, Asia and the Middle East. They present a picture of infinite variety. Each has been subjected to different historical, economic and political influences and has its own individuality. (...).

Dr. Asheim's list of characteristic defects in the library systems of the developing countries (he deplores the almost total absence of public libraries and of service to children, the existence of national libraries with seemingly no function, the lowly status of the librarian in the community and the fact that the book so frequently has primacy over the reader) is reasonably valid for some groups, but not for others. For example, if we consider the library situation in those English-speaking territories which have achieved independence during the past ten years or so, we shall find some at least of his criticisms very wide of the mark. Public library systems have been established in a number of these territories, university libraries are making significant progress, the primacy of the reader over the book is generally accepted and last, but vitally important, librarians everywhere are accorded considerable status in their various communities".(1)

A major obstacle on the way towards coordination of the information activities throughout the world is the growth of "language barriers". The solution of the language barriers problem becomes increasingly complex and urgent from year to year. Whereas early in this century the main languages used for scientific communication were English, German and French, nowadays 50 percent of scientific and technical literary output is in languages unknown to more than half of the scientists of the world. Over the past decade, the proportion of scientific and technical literature published in Asian, African and Latin American countries in Spanish, Portuguese, Hindi, Urdu, Arabic, Suahili and other languages has been rapidly growing.

We believe that the only way of overcoming the language barriers is the solution of automatic translation problems on the basis of an intermediary language. Coordination of national programs for the compilation of bilingual and multilingual dictionaries in science and technology would by very useful. An important prerequisite for a Universal Information System is the compatibility of national scientific and technical information systems available or being developed in different countries. This necessitates the elaboration of international standards for information media, codes and types, transcription and transliteration systems, abbreviations of periodical and serial titles, bibliographical descriptions citations, etc. Also necessary are uniform international rules for the preparation and presentation of primary publications, with special emphasis on providing adequate titles for these publications. It seems to us that it would be most convenient, when publishing a document, to index it by UDC and keywords (descriptors) and to provide it with a brief abstract.

These tasks are to be solved through the joint efforts of international organizations concerned:

United Nations Educational, Scientific and Cultural Organization (Unesco),
International Council of Scientific Unions (ICSU),
International Federation for Documentation (FID),
International Federation for Information Processing (IFIP),
International Federation of Library Associations (IFLA),
International Organisation for Standardization (ISO).

To ensure effective coverage of developing countries by the Universal Information System being developed, wide-scale measures should be taken in these countries to train skilled information workers. This calls, among other things, for the speedy provision of appropriate curricula, programs, and study aids, and scholarships for specialists from developing countries as well as for the creation of a network of training centres both in the advanced and developing countries. This will enable some 80 developing countries to upgrade their information activities thus speeding up their economic and cultural development.

---

1. Unesco Bulletin for Libraries, 1968, vol. 22, N 2, p.90-92.

Thus, we have considered in this lecture several themes which ought to broaden our view of the routes or, as we call them, channels of dissemination of scientific knowledge. You have come to know the fundamentals of organization of information activities, the principal types and functions of the information agencies and special libraries, the stages and routines in the operation of these ancillary institutions of science. This will help you in making use of these institutions in your work. Besides, it will facilitate your understanding of the basic principles of information retrieval, which is the subject of the next lecture.

Checkup questions

1. What are the channels by which scientific information is spread from the generators to the users?

2. What are the main stages in the publishing and printing processes?

3. What are the differences between the scientific information agencies and the special libraries?

4. What principles underlie the scientific information activities and how are they realized?

5. What are the functions of the major information centres in the USSR?

6. What are the missions of the special (research) libraries?

7. What are the main library routines?

Literature

1. Dembowska M. Documentation and Scientific Information. Outline of problems and trends. Translated from Polish. Warsaw, Central Institute for Scientific, Technical and Economic Information, 1968. 154 p.

2. Shera J.H. Of Librarianship, Documentation and Information Science. -- "Unesco Bulletin for Libraries", 1968, vol. 22, N 2, p. 58-65.

3. Kent A. Specialized Information Centers. Washington, Spartan Books; London, Macmillan, 1965. 290 p.

4. Clapp V. The Future of the Research Library. Urbana, University of Illinois, 1964. X, 114 p.

5. Giljarovskij R.S. International Distribution of Catalogue Cards. Present situation and future prospects. Paris, Unesco, 1969. 84 p. (Unesco Manuals for Libraries, 15).

6. Asheim L. Librarianship in the Developing Countries. Urbana, University of Illinois, 1966. VII, 95 p.

5. ESSENTIALS OF INFORMATION RETRIEVAL

The problem of information retrieval is contral to informatics. By now you have learned how scientific data are communicated in the contemporary world, what the sources of scientific information are, what institutions are positioned on the route of scientific information and provide for its transfer from the generators to the users, what difficulties arise in conducting information work by conventional means.

Further upgrading of this work calls for mechanization and automation of information routines. This in turn requires that these routines be formalized. In order to understand the gist of these processes and to identify their common features, they are embraced by the concept of information retrieval. The information retrieval problem can be viewed as the constant accumulation of an ever-growing bulk of scientific information, on the one hand, and the growth and increasing complexity of information needs of the specialists, on the other hand.

In a generalized form the problem can be seen as the need to select for each user from the totality of the available information only those data which are necessary for him at the moment to conduct scientific research or practical work. It is quite evident that no specialist is able to read all existing documents of science in order to select the necessary ones. For this reason, scientific documents are stored in special collections, and the users wishing to extract from these collections the necessary documents use a certain procedure which we call information retrieval.

This procedure can be illustrated by a hypothetical example shown in Fig. 16. Let us assume that we have a collection of ten documents. To eliminate the need for scanning the whole file each time the documents containing the needed data are to be located, the document contents have to be analysed beforehand according to certain characteristics and with a view to the possible questions. If now all the documents are numbered and the characteristics coded by letters, and each document number is linked with its content characteristics, a certain system will evolve which permits formalization of the process of document retrieval. Let us assume further that documents represented by combinations of characteristics C and D or characteristics C and F are to be retrieved. Successive scanning of the columns in which the availability of a characteristic is marked with a dot, permits to establish quickly that documents 1, 3, 6 and 8 meet the above combinations.

This example shows that to implement information retrieval, it is necessary first to select documents for the system which we call an information retrieval system, and to assign each document a number which will serve as its address in the system. Then a multiplicity of characteristics by which these documents will be searched in the system has to be established. This multiplicity of characteristics expressed, according to certain rules, by terms or conventional symbols (numbers) constitutes an information retrieval language.

The subject matter of each document is matched against the terms or numbers of this language in order to establish which of the characteristics they represent are contained in the document. The terms or numbers selected as the result of this matching procedure form the search pattern of a document; its formation is called indexing.

During a search, a user' query is also expressed in terms of the information retrieval language and the search request formulation thus obtained is matched against the documents' search patterns. In case of their complete or partial matching the decision is taken to furnish the appropriate documents to the user. The necessary level of matching is determined by a specified matching criterion.

Basic notions

We shall now define and interpret the basic notions of this section of informatics. Information retrieval is a multiplicity of consecutive operations performed to locate the necessary information or documents containing it, with subsequent retrieval of these documents or their copies. Information retrieval is effected by means of information retrieval systems.

An information retrieval (IR) system is generally formed by an information retrieval language and a matching criterion designed for information searching in a certain information collection. Specific IR systems are realized by means of certain technical facilities, which are nevertheless not included in the abstract notion of an IR system, because the same system can be realized using different means. These technical facilities, various equipments and machinery will be considered in a separate lecture.

The systems divide into document retrieval and data (fact) retrieval systems. Document retrieval systems produce, in response to a query, documents containing the information sought, their copies, or their addresses in the storage. Data retrieval systems are designed to produce facts, e.g. the properties of a particular substance, the structural or molecular formula of a chemical compound, the characteristics of a particular biological species, or the names of those species which possess certain characteristics, etc. The common feature of these systems is that they can retrieve only the information that has been introduced into them beforehand.

At the present time, investigations are under way aimed at creating logical information systems, in which logical processing of information will be possible. Such systems will permit to obtain, in answer to a request, not only the previously introduced information but also new information which has not been explicitly entered into the system. Before such systems can be made operational, however, complex problems must be solved, many of which are of a general scientific consequence. The IR systems under review will form part of more elaborate logical information systems.

An information retrieval language is a major component of an IR system. It is a specialized artificial language designed to express the subject-matter of documents and information requests, in order to locate in an information collection those documents which provide answers to certain questions. Every IR language must meet certain requirements. First of all, each notion in it must be expressed by one and only one word (a certain sequence of symbols) and vice versa, each word must express a single notion. This requirement can be referred to as uniqueness of the vocabulary. Naturally, the IR language vocabulary should not be emotionally coloured, that is, should not reflect the attitudes of the information generator and user to the given notion. Another requirement, that of a formalized grammar, stipulates that any statement formulated in terms of an IR language must admit of a single interpretation. Only when these requirements are satisfied will it be possible to formally match document search patterns and search formulations. It is self-evident that the natural languages, in which one and the same word may have different meanings (homonymy) and one notion may be expressed in different words (synonymy), where the meaning of a sentence is modified by grammatical means or by context, are unfit for information retrieval purposes.

That is why specialized IR languages have long been a standard feature of information retrieval. Conventional kinds of such languages include library and bibliographical classifications (such as those used in a classified catalogue) and alphabetical subject classifications (such as those used in an alphabetical subject catalogue), which we have come across while discussing secondary documents. A feature of these conventional information languages is that the totality of existing knowledge has been distributed in them among the words and phrases in conformity with the possible information requests.

However, such a structure is a source of a certain weakness on the part of the conventional IR languages, as the document contents and user requests depend on the progress of science and technology, on the varying interests and personalities of specialists, i.e. to all intents and purposes, they escape accurate accounting and prediction. The setting up of an IR system, the indexing of documents and their retrieval are separate operations divided in time, sometimes by very large intervals. For this reason a good portion of information contained in documents that were indexed by means of a conventional IR language will be lost in searching. To obviate this disadvantage, new types of IR languages are created in which the words are not grouped in previously established statements (search patterns). This process takes place either during the indexing stage or even during retrieval. Particular IR languages will be the subject of a large portion of the two lectures that follow.

The second component of an IR system, as we have seen, is the matching criterion. A matching criterion is a set of rules according to which the degree of semantic similarity between a document search pattern and a search request formulation is established in an IR system and the decision is taken as to whether to retrieve or not to retrieve the given document in response to a query. Special attention should be paid to the notion of semantic similarity contained in this definition; it has been termed relevance. At the present time this semantic similarity (relevance) cannot be yet established formally and hence, automatically. Therefore, the relevance judgments, normally made by the user himself, will be necessarily subjective.

The situation is further complicated by the fact that requests are not adequate representations of the specialists' information needs. Information needs are dynamic in character and will continually change with the receipt of new information. When formulating a search request a specialist is not always aware of this actual information need. Getting to know the contents of the documents retrieved may change the specialists' idea of his actual need, which will immediately affect his request formulations. Thus a sort of user - system dialogue takes place in which the operation of the whole system is optimized through a succession of iterations.

Something like this is observed in everyday life. When a person applies to a consultant, the latter will supply an answer not before he identifies, by posing a number of questions, the actual information need of the applicant. The latter may sometimes gain full realization of his needs only following a talk with the consultant. A strong side of the conventional IR systems, e.g. catalogues of libraries or bibliographic indexes, is that the feed-back to the user is a built-in feature of their structure. Automated document retrieval systems should also be designed in such a way as to enable the searcher to investigate the search file, modifying his search request formulations in conformity with the intermediate retrieval results. In other words, they must be designed as systems of the "man-machine" class.

Another way of overcoming this difficulty is the perfection of the IR languages and matching criteria. For instance, the matching criterion in a Soviet IR system, named "Pusto-Nepusto" (or "Empty-Non-empty"), provides retrieval of documents not only when a match of a term or terms occurs in a document search pattern and a search request; if desired by the user, the system can retrieve a document if the terms in its search pattern occupy a higher or lower position in a logical hierarchy than at least one term in a request formulation. Thus, in response to a query concerning documents the subject of which is the properties of some liquids, the system can provide documents dealing with the boiling point of water, or vice versa.

IR system efficiency

It has been noted before that a document search pattern recognizes only the main subject treated in the document. Therefore, such a method of document identification cannot ensure the complete recall of all pertinent information in each particular case. On the other hand, part of the retrieved documents may not contain pertinent information at all. These documents constitute the so-called false drops, or noise, and reduce the precision of information searching. The incompleteness and imprecision in information retrieval are a kind of price that has to be paid for facilitating the searching procedure. The price is lower with the more advanced IR languages and matching criteria and, consequently, with the more elaborate searching strategies.

This reasoning brings us to an understanding of the methods of evaluation of document retrieval systems. Operational efficiency of an IR system (economic efficiency will not be discussed here) can be defined as the measure of the system's ability to discharge those functions for which it is designed. The function of a document retrieval system is to pull out, in answer to a request, from an information file relevant documents, i.e. documents that are semantically related to the query. As a rule, in experiments comparing the efficiency of several IR systems the documents are assessed for relevence by a team of judges. In practical operation of IR systems, the relevance judgments are made by the user himself, who assesses a document content in terms of his actual information needs. It follows that all evaluation studies are formal in character and their significance lies mainly with comparisons of different types of IR systems.

The most widely used measures of IR system effectiveness are recall and precision. Recall can be defined as the ratio of the number of relevant documents retrieved to the total number of relevant documents in the information file. The precision refers to the ratio of the number of relevant documents retrieved to the total number of (both relevant and non-relevant) documents retrieved. Both measures are usually expressed as percentages. The upper part of Fig. 17 illustrates these definitions by a contingency table of relevance and recall.

As has been demonstrated in the experiments of the British documentalist C. Cleverdon, popularly known as the Cranfield Project, there is an inverse relationship, although not strictly formal, between recall and precision values. An increase in precision will bring down recall. This relationship can be traced in the diagram also contained in Fig. 17. Therefore, increasing the complexity of an IR language and matching criterion, and thereby the searching procedure generally, with the aim of reducing noise will result in the loss of many relevant documents. If a 100-percent precision is sought for, i.e. only relevant documents are wanted in the search output, the recall will approximate zero. And the other way round if the aim is a 100-percent recall, i.e. retrieval of all relevant documents in the collection, the precision will approximate zero on account of the increased noise.

The choice of the recall-precision ratio is of major importance in developing specific IR systems. Most systems have a recall of from 70% to 90%, and precision, 8% to 20%. These values can be held to be satisfactory for meeting information requests of scientists and engineers, because with a certain information redundancy of the contents of primary scientific documents, the loss of a small proportion of relevant documents will not be felt, and the possibility of iterative searching will help to overcome the adverse effects of noise. When the recall reaches 90%, its further increase will be attended by a sharp drop in precision. It would not be an exaggeration to say that the provision in the search output of the last 10% of relevant documents will demand a higher price than all the previous output. The data cited indicate that there are limits to the efficiency of any IR system, and that both information losses and noise have to be accepted in their operation.

General framework of an IR system

Although the notions of information retrieval and an information retrieval system had their origins several dozen years ago, the corresponding processes and facilities have been in use in scientific work for hundreds and thousands of years. It was the achievement of informatics that the operations of a multitude of empirically evolved facilities for information retrieval (reference books, catalogues, bibliographical indexes, libraries, etc.) could be fitted into a common framework. The general layout of an IR system is shown in Fig. 18.

We shall analyse its structure in terms of a document retrieval system. You will recall that such a system is designed to retrieve scientific documents in answer to the information requests of the users. Therefore, when an IR system is set up, it should be provided with an initial input of documents, and the new emerging documents should be added to it in the future. This process is shown symbolically in the left-hand part of the scheme. The documents come to the Input Converter (IC) where they are subjected to a number of operations. Each document undergoes an analytico-synthetic processing yielding its search pattern. The search pattern contains a brief formal description of the given document which distinguishes it from all the other documents (bibliographical entry), as well as a concise statement of its subject matter (notation, terms, annotation, abstract).

In addition, each document is assigned an address (usually a number), by which it can be located later. The documents themselves can be converted into a different form (miniaturized or machine-readable). This completes the conversion of the documents at input into the system, and the documents (in the original or converted form) with their storage addresses proceed to the Passive Storage (PS). They are kept there until requested by the information users.

The search patterns recorded on some material medium (catalogue cards, punched cards, microfilm, magnetic tape) together with the respective document addressed, are separated from the documents and forwarded into the Active Storage (AS). There they are arranged in an order facilitating their matching against every incoming request.

The input of information requests into the system is symbolically shown in the right-hand portion of the scheme. First of all, they go into the Input Device (ID) where they are converted from the searcher's natural language into the information retrieval language used in this system. After the conversion they emerge as search request formulations, suitable for matching against the search patterns stated in the same information retrieval language. The request formulation proceeds into the Resolver (R) through which are passed the search patterns of the documents from the AS (that is why it is called "active"). There, in the Resolver, the search request formulation is matched against the search patterns. In case of a match (complete or partial, depending on the matching criterion applied), the R gives an instruction to the PS to retrieve the corresponding document by its address. Between the PS and the user there is usually an Output Device (OD) which provides for the return of the documents into the PS, and also converts into the original form the documents which have been stored as microreproductions or as machine-readable records. The furnishing of the wanted documents or of their hard copies completes the operation cycle which is started by a user request. Such are in very broad outline the structure and operations of any information retrieval system.

Examples of specific IR systems operation

This can be illustrated using the already familiar example of library operation. The Input Converter (IC) in a library is the cataloguing (processing) department. It is there that a record is made on catalogue cards of the search pattern of a document (bibliographical entry, classification numbers, subject headings) and its address (call mark), which is also inscribed on special labels pasted onto the documents (books or journals). The PS is the stacks where the books are shelved according to their call marks, and the AS is the catalogues, where they are searched. The bibliographers in the bibliographic reference department and the catalogue-room attendants serve as the Input Device (ID), and the reading-room or counter attendants, as the Output Device (OD). The Resolver (R) in this case is simulated by the intellectual efforts of the reader who scans the cards in the library catalogue, deducing from the entries which books may be of use to him for solving the problems of his concern.

This example can help to spot some faults of the conventional library as an information retrieval system, in addition to its general shortcomings which were discussed in the previous lecture. They consist primarily in the principles of organization of the Passive and Active Storages.

The PS in the library takes the form of tier stacks, where the document originals are arranged on the shelves. These documents occupy a great amount of space, their issue, shelving arrangement and rearrangement are very labour-consuming and do not lend themselves readily to machanization. In consequence, the greatest difficulties in the work of the modern libraries are those presented by the lack of suitable storage room for the stocks. Besides, the lending of literature can lead to situations where requests for the literature of constant demand cannot be met because the books sought are not in the stacks, having been issued to other readers. These difficulties can be overcome only if the documents in the PS are stored as microreproductions or machine-readable records and the readers are issued only true-size hard copies of these documents.

The active storage, realized in the library in the shape of card catalogues, also suffers from serious disadvantages. Because the cards in a catalogue are filed in a lineal sequence, any one catalogue can interpret a book collection in one aspect only. We have learned already that the librarians make do with three such aspects: author, subject and classified. Within an aspect a separate card is made for each characteristic for which a search may be conducted.

Thus, increasing the number of aspects and characteristics by which a reader can search his literature will result in more complicated and cumbersome systems of catalogues. This, in turn, increases the labour costs of catalogue maintenance and hampers their use by the readers. For this reason the librarians are obliged to restrict the number of catalogues in libraries, as well as the amount of duplication of bibliographical entries. This, of course, detracts from the search capabilities of the library as an information retrieval system.

This disadvantage of the library catalogues as active storages springs from their inherent organization, which requires that each document be provided with a search pattern on a separate information carrier - a catalogue card. A card duplicate with the given search pattern must be reserved for each search characteristic. A card duplication ratio of 1.5 for each catalogue has come to be accepted in library practice. This means that within each aspect a document will have, on the average, one or two search characteristics. But even with this restriction, a library reflecting its stock in three catalogues will have to keep four or five cards per document.

Such an organization of AS is called serial. It presents another inconvenience, namely, that separate searches are necessary for each aspect or characteristic, and multi-aspect searching is difficult. Most mechanized and automated IR systems use an inverted organization of the AS. Essentially it means that a separate information carrier (e.g., a card) is assigned a certain search characteristic or aspect. The addresses of all documents whose search pattern contains this characteristic or aspect are recorded on this carrier.

The inverted scheme has several advantages in comparison with the serial one; it is more compact, it permits multi-aspect searching, and by the feed-back to the user it permits adjusting the search strategy. Adding or deleting certain search characteristics in the request formulation will reduce or expand the search output. These problems will be treated in more detail later on, while discussing descriptor IR systems and technical means of their implementation.

It should be mentioned in conclusion that the one important advantage of the conventional library IR systems over mechanized and automated systems is that in the former the functions of the Resolver are performed by the reader himself, who during his searches in catalogues can modify his requests and the intuitively applied matching criterion. This to some extent compensates for the shortcomings of the organization of the AS and makes for higher efficiency of the library IR system as a whole.

In this lecture we have considered the essentials of information retrieval, defined the main concepts of this part of informatics, made acquaintance with the methods of assessment of IR system efficiency, with the schematic diagram of an IR system and the interactions of its component parts, discussed the advantages and shortcomings of the different methods of organization of active storage. All this information forms the necessary background for a deeper understanding of the structure and principles of utilization of the existing conventional and automatic IR systems, which will be dealt with in the next lectures.

Checkup questions

1. What is information retrieval and what are its basic principles?
2. What component parts constitute an information retrieval system and what are their functions?
3. What is the gist of the relevance problem?
4. How is the efficiency of an IR system determined?
5. What is the purpose and structure of the schematic diagram of an IR system and how do its components interact?
6. Describe the operations of a library and the use of an abstract journal in terms of IR systems.
7. What is the advantage of the inverted organization of AS over the serial one?

Literature

1. Kent, A. Textbook on Mechanized Information Retrieval. New York, Wiley, 1962.
2. Vickery, B.C. On Retrieval System Theory. 2nd ed. London, Butterworths, 1965, XII, 191 p.
3. Fairthorne R.A. Towards Information Retrieval. London, Butterworths, 1-61. XXIII, 211 p.
4. Sharp, J.R. Some fundamentals of Information Retrieval. London, Deutsch, 1965. 224 p.

6. CONVENTIONAL INFORMATION RETRIEVAL SYSTEMS

Information and bibliographic publications as well as catalogues of libraries and bibliographical indexes, which were the subject of our third lecture, may serve as particular examples of conventional IR systems or their separate parts.

In this lecture we revert to them in order to consider the principles of their design and operation in terms of information retrieval.

We know already that the basic types of conventional document retrieval systems are the author, alphatetical subject and classified systems. Since in conventional IR systems the matching criterion is not explicitly stated (it is applied by the searcher intuitively), their sole component is information retrieval language. In the author systems the IR language is formed by the rules for bibliographical description and alphabetization; in the alphabetical subject systems, by lists of subject headings and methods of subject indexing, and in the classified systems, by hierarchical or faceted classifications and classification methods. These conventional-type IR languages will be the subject of this lecture.

Author systems

Author systems originated many centuries ago as bio-bibliographical lists and were designed mainly for the identification, first, of different manuscripts, and subsequently, of different editions of the same work. The turning-point in their development and use was the mid-19th century when they enjoyed widespread acceptance as author catalogues in large public libraries. It was then that the theory of cataloguing and author catalogue arrangement began to develop, giving birth, by the 20th Century, to the various national cataloguing codes, and in our day, to the internationally approved principles of descriptive cataloguing.

The main idea behind the rules for bibliographical description is to isolate in every type of work a bibliographic particular (e.g., names of the authors, the title, the sponsoring institution) which would best characterize the given type. These basic data, extracted from a work according to certain rules and recorded in a symbolic form, constitute what is called an index entry; the totality of these data make up the vocabulary of an author system IR language. The other data, also extracted and recorded according to special rules, constitute the text of the entry and form an integral part of the search pattern of any document in an IR system. The entries are arranged alphabetically by headings in conformity with strictly defined rules which serve as the grammar of the given IR language.

The need for elaborate rules for bibliographical description follows from the IR system requirements which we formulated in the previous lecture, namely, the uniqueness of the vocabulary and the formality of the grammar. But the bibliographical data are, unfortunately, far from being unique and their manner of presentation varies in different publications. For this reason the descriptive cataloguing methods rigidly regulate the sources, kinds and elements of an entry, as well as its language and spelling. In principle, the source of description is the document as a whole, but primarily its title or front-page which is characterized in the language of the original but in modern spelling. Abbreviations of certain words, prescribed by a special list, are used in the text of the entry.

The bibliographical entry contains the following particulars: (1) Author(s), or the heading. (2) Title. (3) Sub-heading data - the data explaining and qualifying the theme and content of the document, its nature and purpose; data concerning the approval or endorsement of official documents; data on the persons participating in the compilation of the document or its preparation for publication; data on the edition or the official character of the publication. (4) Imprint - place of publication, publishers, and year of publication. (5) Collation - number of pages, and the presence of illustrations. (6) Information before the title - the name of the institution which sponsored the publications; the title of the series in which the work appears, and its serial number. (7) Bibliographical notes - data on the authors given in the publication (if they disagree with the data accepted for the entry, or if the work has more than two authors); data on the variant readings of the title, on the availability of a bibliography in the document, on publication defects. Fig. 19 gives a breakdown of bibliographical data on the title-page of a publication and in its entry.

Different approaches to descriptive cataloguing arise in connection with the authorship, number of volumes, frequency of their appearance and the polygraphic autonomity of a work. Authorship is the most individualized attribute for some of the more widespread publication types. Works by one, two and three authors are entered under the surname and forename (initials) of the first author. An added entry is made for the second and third co-author (or they are listed in the alphabetical index of names), so as to reflect the given work in the common alphabetical sequence of their works. The names of these authors and presented in the headings in a unified form, in order to ensure the unique location of the works of the same author in the common alphabetical listing.

The authors are called compilers in such publications as dictionaries, guides, handbooks, study aids and guidance manuals, reading books, antologies, technical and scientific information materials, advanced methods exchange materials, and fundamental bibliographical indexes. All these publication types are indexed under their compilers, who have done a piece of full-fledged research.

Works by four and more authors, collections of works of different authors, syllabi, procedure manuals, statistical handbooks and similar publications are entered under the title, even where the compiler or editor are indicated. (In some countries, e.g. in English-speaking countries, such publications are indexed under compiler or editor). Added entries are made for the first author, if a work of four or more authors is indexed, and for the compiler or editor, if a collection is indexed, which is a great help to those users who will memorize a book by such names rather than by its title.

Official publications are issued on behalf of institutions, organizations or agencies as their documents for which they bear full responsibility and which are the end product of their activities. These include statutes, regulations, manifestos, appeals, reports, plans, decrees, resolutions, instructions and other similar materials. Their titles usually begin with these standard words and contain the name of an institution or organization, e.g., "The Charter of the Communist Party of the Soviet Union". The salient feature of an official publication, which determines information requests for it, is the name of the institution which put it out. For this reason official publications are commonly indexed under the name of the institution, which in this case appears as a corporate author (not to be confused with composite authors, which are several persons joined in the writing of a work). The same indexing procedure is used with many non-official publications, such as "Transactions", "Proceedings" or "Papers" of institutions of science, congresses, symposia, if their title begins with or is confined to these standard words, and contains as an integral part the name of an institution or organization. For example, "Transactions of the Krupskaya Institute of Culture in Leningrad".

According to the number of volumes (issues) and their frequency of appearance publications are divided into multi-volume and serial books, and periodicals. The typographic make-up of a multi-volume publication or a series is such that many bibliographical data will recur in each of its volumes or issues. For such publications a series entry is made, the general part of which gives all the recurring details, and the specification lists only those details which distinguish one volume from another. This saves a lot of entry space and indexing time. It is important though, that not only the similarities but also the differences between multi-volume books and series should be realized. A multi-volume publication is limited by its content to a certain number of volumes, while a series comprises quite independent works united only by the commonness or similarity of their subject matter. Therefore, the series entry for a multi-volume book is its main - and normally its only - entry, citing full bibliographic details. For serial publications, on the contrary, main entries will be the entries for individual issues, which ought always to be made. Here the series entry is an auxiliary one; it is very brief and only made in those cases when a series as a whole can be of scientific interest.

Periodical publications, which are taken to include serials, journals and newspapers, are also indexed by means of a series entry, usually under the immediate title. The only exception to the general rule is constituted by the serials which occupy an intermediate position between the book series and the journals. Occasionally, they are entered under a corporate author and, in addition, separate entries are made for the individual volumes or issues which may have an author or be collections of works having a title of their own.

To index a part of a publication or work which is in itself an independent work (e.g. a paper in a collection or journal), an analytical entry is made under the author and title, as with the other independent works. However, in place of imprint, details of the publication in which the work is included are indicated. These details may vary according to whether the publication in question is a collection, serial, journal or newspaper. An entirely different matter is making an analytical entry for a part of a work, such as a chapter or section of a book, which cannot be viewed autonomously, out of the context of the work to which it belongs. Therefore, chapters or sections are entered under the author and title of the corresponding work. But instead of the collation such an analytical entry cites the number and title of the chapter or section indexed, as well as the pages it occupies in the book.

The same procedure applies to reviews, or critical comments on books, which usually appear in journals, serials or collections. What the reader wants to know is primarily whether there is a review for this or that book, and often he will not know who wrote the review, how it is titled and where it is published, for this reason reviews are entered under the author and title of the book reviewed. The entry lists all the particulars which are usual in book entries up till collation, instead of which details of the review and of the publication in which it appears are given. The most important reviews may be also indexed as separate works. The above rules for description of the main publication types are schematically displayed in Fig. 20, with illustrative examples shown in Fig. 21.

Rules for alphabetical filing of entries, which form the grammar of this conventional IR language, specify the various methods of such filing within one and within several natural languages. There are two commonly used methods of alphabetization of words, names or sentences, which are both based on the order of letters in the first words. By the first method ("word-by-word"), entries are filed in the sequence of words, by the second ("letter-by-letter") strictly in the sequence of letters of the alphabet, without consideration for the division of sentences into words, e.g.

"Word-by-word"	"Letter-by-letter"
New England	Newark
New wives for old	New England
Newark	Newman
Newman	New wives for old

Most alphabetical systems employ the first method as the more common-sense and less formal. It applies not only to prepositions and other function words (articles, particles), but also to complex and compound words, including forenames and surnames. In some systems this rule is made still more complicated: if the first words coincide, it is not the alphabetic sequence of the words that follow which is taken into consideration but the heading category: entries are listed first under the names of the authors, then under the names of institutions, and finally, under the titles beginning with the same word. Since this facilitates locating entries in long alphabetical sequences, it is sometimes practised to list these three heading categories in three different alphabetic sequences.

There is another range of problems of alphabetization of index entries, which is connected with the fact that different languages use different alphabets and systems of writing. Most of the nationalities of the USSR, for instance, make use of Russian letters, many nations in Europe, America and Africa use Roman letters, some nations of Asia employ national systems of writing, e.g. the Arabic, Red Indian, Chinese hieroglyphic etc. Moreover, different languages using the same letters may have different alphabets. Thus, the alphabet of the Azerbaijan language differs from that of the Russian language, and the English alphabet differs from the Polish one. Because of that, entries in languages which have different letters and alphabets have to be placed in different alphabetic sequences. If an occasion demands a common alphabetic sequence for entries in different languages within the same system of letters, the basis is provided by some artificial alphabet that usually ignores auxiliary letters and diacritics.

In conclusion it should be noted that until quite recently the alphabetic systems had been the only means of identification of publications, of searching for specific titles in IR systems, and of searching for works by specific personal and corporate authors. Excessive labour costs incurred due to the existing cataloguing rules and their insufficient standardization on the international level should be pointed to as the principal faults with these systems. An attempt to rectify these inadequacies was the recent proposal to use standard numbers for books and serials as their identifiers in information work. The main idea of the proposal is to assign to every new book and every serial title a unique number associated with the respective entry in a national bibliography. Implementation of this proposal would appreciably facilitate the input of documents in automated IR systems and the exchange of bibliographical data both on a national and international scale.

Subject systems

We have already seen, as in the case of permuted-title and citation indexes, that a listing of some particulars of a bibliographic entry can, under certain conditions, provide a subject searching capability. This fact has long been utilized in subject (or alphabetical subject) systems. Examples are subject catalogues and alphabetical subject indexes which were discussed in the second lecture.

Subject cataloguing is a special kind of document classification which is concerned with the concise formulation of document contents by means of one or several standard words called subject headings. These are arranged alphabetically and followed by either bibliographical entries, as in subject catalogues, or classification numbers, bibliographic entry numbers or book page numbers, as in alphabetical subject indexes.

The task of subject cataloguing is to identify the subject which is the main theme in a document, as well as to establish the main aspect of its characterization and its relationships with other subject. The different aspects of the characterization of a subject are its history (origins and development), structure, composition, properties, state, purpose, interaction with other subjects, investigation, evaluation etc. In subject cataloguing, documents are grouped by subjects not belonging exclusively to any one area of knowledge. This makes it possible to form different complexes ("subjects") such as a personality, human activity, geographical notion, natural phenomenon, social phenomenon, material, property, etc. Thus, the subject heading "X-rays" will bring together the documents which are concerned with the nature of X-rays (physics), their use in medical diagnostics (medicine), design of X-ray apparatus

(X-ray engineering). As subject headings are listed in a common alphabetical sequence, it is not necessary, while searching for a document on a specific subject, to know the place of this subject in the general systematics of scientific knowledge. Subject headings of quite different content will be listed side by side in alphabetical subject systems, e.g.:

Apricots
Archangelsk
Aristotles
Automobile

On the other hand, alphabetical arrangement joins together words of the same root which often prove to be associated semantically; this will help to create in a subject system a broad thematic complex, e.g.:

Atom
Atomic bombs
Atomic electric power station
Atomic energy
Atomic ice-breakers
Atomic nucleus
Atomic spectra
Atomic thermal capacity

A subject heading may be selected from a special previously established list or it may be formulated by the indexer; in either case it recognizes only the main aspects of a document content, which in practice is usually covered by a very few headings. The major problems faced in building up the IR language of a subject system are the use of "generic" and "specific" headings, the inverted word order in the formulation of subject headings, and the use of sub-headings. Without going deeply into the methodology of subject classification, solutions for each of these problems can be illustrated by a few examples.

A subject heading must be most specific in defining the subject of a document. Thus, a document dealing with motor-car carburettors should be entered under "Carburettors" and not under "Internal combustion engines". In this case the subject heading is "specific", or rather, adequate, i.e. corresponding to the main theme of the document. If, in anticipation of the possible requests of the system users, it is thought desirable to enter the given document under a heading expressing a more general idea (in our example "Internal combustion engines"), such a heading is taken to be "generic", or more correctly; generalizing.

In order to join together related subject materials in a subject system, the words in a heading can be reordered (inversion). Thus, adjectives are often placed after a noun, e.g.:

Microbiology - veterinary
Microbiology - medical
Microbiology - industrial
Microbiology - agricultural

In some cases, however, inversion is out of place; as when an adjective defines an important attribute of a subject ("Computational technology", "Electronic valves"), or in the names of historical events and other fixed phrases ("Roman Empire"). In order to divide the materials within a subject heading subheadings are used, e.g. topical ("Water - Analysis, Evaporation, Purification, Chlorination"), geographical ("Animal Breeding - USSR, Great Britain, France, USA"), form ("Standards - Bibliography, Classification Schedules, Indexes"), chronological ("Periodicals - History - 17th Century, 18th Century, 19th Century"). A special position is occupied by the standard or common subheadings, e.g. those used in the subject cataloguing of the literature on materials ("Concrete - Analysis, Defects, Substitutes, Testing").

Subject systems are most effective in searching for narrowly defined subjects and far less useful for broad thematic searches in a particular area of knowledge. This is their peculiarity which should not be viewed either as an advantage or disadvantage: it merely defines the conditions of their application. On the other hand, subject systems have a number of unquestionable advantages. They are simple in operation and require no preliminary training on the part of the user; introduction of new subject headings does not involve any changes in the existing headings and re-indexing of the documents already in the system. Growth in the number of subject headings and documents in a system does not create any special difficulties either for the indexers or for the users. Subject systems are easily implemented as either manual card catalogues and card files or as bibliographical lists.

But subject systems also suffer from a number of shortcomings. They will retrieve only those documents whose main subjects match document requests, and will not recognize the other subjects discussed in scientific documents. (This, however, is the common inadequacy of conventional IR systems oriented to manual handling). Another shortcoming which is equally common to all conventional IR systems is the difficulty of multi-aspect searches, since a subject system will not permit to obtain simultaneously details of several documents entered under different headings, that is to say, it is impossible to conduct a search using any combination of characteristics, or their logical product. The retrieval languages of subject systems employ the vocabulary of some natural language and are therefore unsuitable for international use. Finally, in comparison with other conventional IR systems, the setting up and maintenance of a subject systems, is a very laborious task for highly skilled personnel.

Hierarchical classifications

Classification is the grouping of subjects or relations in classes according to a common property which is inherent in all the subjects of the given kind and distinguishes them from the subjects of the other kinds, the grouping being done in such a way as to put each class in the system in a definite, fixed position with respect to the other classes. The property according to which classification is done is called a characteristic of division.

The classification process obeys some formal rules of logic: subjects or relations may be divided according to only one characteristic of division at a time, and the resulting classes and subclasses should be mutually exclusive, the division into classes should be balanced and continuous, without leaps. Classifications in which every subclass has only one class that immediately precedes it (relations of strong hierarchy), and all subclasses are subordinated to only one more general class (collateral subordination relations) are called hierarchical. By way of visual demonstration of a hierarchical classification, Fig. 22 shows its graph, that is a diagram consisting of points (apices) and the lines (edges) that connect them, as well as an Euler-Venn diagram indicating the relationships between the classification divisions.

Classification of documentary information, which is one of the most important IR language types in conventional systems consists in the grouping of scientific documents by parts of knowledge according to their content. Such document classifications are more or less associated with classifications of sciences, as their main classes will broadly accord with the particular fields of knowledge, and further division of these classes will correspond to the structure of these fields of knowledge.

However, these classifications, which have been termed library classifications, are not identical with classifications of sciences. Library classifications must be built according to some formal rules of logic, because only under this condition can they serve as IR languages and provide unique identifications to the documents. But rules of logic are inapplicable to classification of sciences, since there are no clear-cut boundaries between the different sciences. The other distinctions also spring from the strictly practical character of library classification and the peculiarities of the subjects classified. Aside from the division by document content, a library classification must include division by types of publications (books, periodicals, special technical publication types), by their purpose (scientific, popular science, instructional), by language, etc.

Schemes of library classification are generally published as schedules consisting of two parts: the main tables and the tables of auxiliaries or common subdivisions. Main tables list all fields of knowledge and their sections in a logical order, each level of division accomplished according to only one characteristic. Each heading has a number of subordinate headings forming a certain hierarchy. Because of these characteristic traits the conventional library schemes are sometimes called linear-hierarchical. Tables of auxiliaries or listings of common subdivisions display the recurring characteristics of different subjects, e.g. the characteristic of place and time to which a document content refers, or that of type of publication.

Each division of a classification is assigned a conventional symbol called a number. The totality of classification numbers forms its notation which may be numerical, alphabetical, or mixed. The merits of a numerical notation spring from the fact that a numerical sequence is more obvious and familiar than an alphabetic one; that any combinations of digits are easily pronounceable as numbers or figures; and that the Arabic numerals are understood by all nations whatever their spoken or written language.

At the same time numerical notations have a serious drawback, namely, a limited base. Because there are only ten digits (from 0 to 9), in order to designate complex or very specific notions very lengthy numbers comprising many digits have to be built. A solution is frequently sought in the use of mixed numbers which include letters as well as numerals.

From the standpoint of structure, notations may be non-structural and structural or hierarchical. Non-structural notations make use of the serial numbers of classification subdivisions in a common numerical sequence. Such notations have low mnemonic power, do not reflect the hierarcy of classification subclasses, and will not admit of further division of these subclasses. For this reason most of the modern library classifications employ structural notations which may be either numerical, alphabetic or mixed. Structural notations represent the conceptual structure of a classification, since every class is designated by one symbol, and all divisions of this class, by two symbols, of which the first stands for the class, and the second, for the corresponding subclass. Structural notation will permit an indefinitely long process of detailing of a classification scheme.

Universal Decimal Classification

Library classifications have been with us from very ancient times, and this is hardly surprising, as the need for such classifications had arisen simultaneously with the emergence of the written records. Nowadays, dozens and hundreds of various classifications are in use.

The Universal Decimal Classification is admittedly the most significant of the classification schemes that were developed around the turn of this century, as well as the one which is most extensively used in all countries of the world. It owes its origins to the work of the two eminent Belgian bibliographers, Paul Otlet (1868-1944) and Henry La Fontain (1854-1943). The international conference of bibliographers which was convened in 1895 through their efforts, formed two organizations: Bureau International de Bibliographie and Institut International de Bibliographie which were later to unite under the other name. The Institut International de Bibliographie which had its headquarters in Brussels was entrusted with the task of compiling an Universal Bibliographical Repertory, i.e. a bibliography that was to cover the literature in all fields of knowledge published in all countries in all languages since the ancient times to our day. Although this evidently unfeasible task has never been realized, it required for its fulfilment the provision of a deep classification scheme that encompassed all parts of knowledge (that is a comprehensive classification), and which could also be used on an international scale.

The basis for the new classification scheme, developed by a group of experts headed by P. Otlet, was furnished by the Decimal Classification of Melvil Dewey, which was considerably reworked. The new decimal classification version was brought out in French in separate issues since 1897. In 1905 it appeared as a complete work, and in 1907 reprinted under the title "Manuel du repertoire bibliographique universel".(1) The present-day editions of this version which are

---

1. Manuel du repertoire bibliographique universel. Bruxelles, 1907.

published by the International Federation for Documentation, the successor to the Institut International de Bibliographie, and also in the different countries in different languages, have been called the Universal Decimal Classification, or UDC.

The UDC comprises the main tables, the auxiliary tables and the alphabetical index. The main tables contain the numbers by which the documents are systematized by their contents. Each concept presented in the main tables must have a certain UDC number. The sum total of human knowledge is divided into 10 main divisions, shown in Figs 23 and 24. Each successive digit added to the designation of a main division does not change its general content but merely qualifies it. Such a method of building UDC numbers makes it possible to divide any general idea into narrow specialized topics. The greater the depth of division the longer will be the notation. Thus, "compressive strength of grounds" in construction engineering is denoted by 624.131.439.4. To facilitate the reading of the notation, every three digits are followed by a full-stop. The numbers are pronounced as a sequence of integers, e.g. six hundred and twenty-four, one hundred and thirty-nine, four. Fig. 25 instances an example of interpretation of a UDC number, 536.23 Heat Conductivity of Gases. UDC numbers are filed, within one division, according to the principle from the genus to the species. Their order depends only on the sequence of figures in each division and not on the length of a number.

Aside from the main classification numbers the main tables list the special (analytical) subdivisions which represent the characteristics and properties typical of a limited range of notions and are limited to use in the section which they immediately precede. They are joined to the main numbers by a hyphen, if applicable in many subdivisions of a given section, or by means of .0 (point 0), if applicable only in the subdivision concerned.

The auxiliary tables list the common auxiliaries which represent the recurring characteristics according to which documents can be divided in all the parts of knowledge and which are applicable in all sections of UDC, as well as special signs which serve to join several main UDC numbers. The common auxiliaries may be of language, of form, of place, of time, of race and nationality, of viewpoint. They are used in all sections of the scheme in the same unchanging meaning. If occasion demands, letters of the Roman or Cyrillic alphabets, separate words or names (e.g. 92 Einstein, for a biography of Einstein) can be attached to UDC numbers.

For a more exhaustive and accurate representation of the subject-matter of the documents classified, the UDC employs, in addition to the main numbers and auxiliaries, the joining of several numbers with the help of different symbols. The common auxiliaries and the connective symbols are shown in Fig. 24. The addition (plus) sign is used in those cases when a document deals with several separate concepts of equal value. The extension (stroke) sign is used to join the first and last of a series of consecutive numbers denoting closely related concepts. The relation (colon) sign permits to classify documents dealing with concepts that are internally related. The synthesis (apostrophe) sign is used with documents on chemical compounds and alloys.

The task of classifying by the UDC is made much easier by the availability of the alphabetical subject index to the tables. Having established the primary subject-matter of a document you will have to look up in the index the division or subdivision that corresponds to the document content. Then you should learn the makeup of the relevent section, select the most appropriate heading and write down its number. In establishing the heading it is not necessary to follow the last level of division in the tables used. You should be guided here by the availability of materials on this subject and the prospects of their growth. The use of too general numbers, though, is undesirable, because then you will sacrify the depth of classificatory analysis of your document.

In classifying documents in the mathematical, natural and the applied sciences (UDC divisions 5 and 6) the most important thing to do is to delimit their topics. It should be remembered that the natural sciences study the laws of nature, while the applied sciences use their findings for practical purposes. It is known, for instance, that X-rays are studied by physics, while the different aspects of their application, by technology and medicine. Hence, the documents which describe the direct observations or results of laboratory studies of X-rays are entered under 537.531 X-rays and Gamma-ray. Non-Corpuscular Rays in Discharges; documents on the design and manufacture of the various X-ray apparatus, under 621.386 X-ray Tubes and Accessories, and the materials on the diagnostic application of these rays, under a subdivision of Medicine, 616-073.7. The use of other subdivisions is also possible here. For example, if a document deals with the subject of films for X-ray photography it should be indexed by 771.531.34 in Photography and Cinematography.

The UDC has the following specific features: (1) coverage of all fields of knowledge (the recent UDC editions contain in excess of 100,000 subject headings); (2) the decimal principle of division which permits unlimited division of subclasses without doing violence to the basic structure of the scheme; (3) the use of exclusively numerical notation which is relatively easy to memorize and equally comprehensible to specialists speaking different languages; (4) the availability of an elaborate system of auxiliaries; (5) the use of the principle of synthetic notation; (6) the possibility of ordering any number of documents to any level of division.

The primary distinctive features of the UDC are its elaborate system of auxiliaries and the synthetic notation. It was precisely because of these features that the emergence of the UDC meant in a way a break-through in library classification. The former "enumerative" schemes with previously established headings and ready-made numbers had been ousted by a new flexible classification in which the necessary headings are formed during the classification process either by combining numbers with auxiliaries or joining together two or more numbers. The existence of a well organized international system for keeping the UDC up-to-date is its unquestionable merit and places it in an advantageous position with request to any of the present-day classifications. As is natural, the institution of such a system takes much time and its operation presents great difficulties and entails much spending. But it is only through such a system that the development and perfection of a classification is possible.

At the same time it should be noted that the division of the whole universe of human knowledge into mere ten classes and the order of these classes in the UDC do not accord with the present level of the development of science. Despite this and some other defects the UDC has become widespread and found application in thousands of institutions in 50 countries of the globe. Presently the size of the complete UDC schedules is approximately 500 author's sheets, which occupy some 10 volumes each 300-500 pages large.

Concluding, it would not be out of place to enumerate the principal merits of UDC which are: universality, international usability, the decimal system of notation, the number-building principle and a smoothly functioning body for keeping it up-to-date. At the present stage of development of the classificatory theory and practice the UDC remains the only internationally accepted universal system which enables to reveal the contents of reference collections in sufficient detail, ensures speedy information retrieval, and makes for closer international co-operation. These merits of the UDC point with certainty to the profitability of using it in scientific information agencies and in the special libraries for classification of the literature in the natural sciences and technology. The Council of Ministers of the USSR has decreed on 11 November 1962 the compulsory use of UDC by the information bodies and technical libraries for classification of literature in the natural and technical sciences.

Decimal classification has been in use in the USSR since 1921. The public libraries still utilize the different versions of the library classification schedules which are based on a modification of UDC made by Soviet bibliographer L.N. Tropovskiy. Work on a new Soviet scheme of classification for the research libraries, based on the Marxist-Leninist classification of sciences, had been going on for many years; it appears in separate issues since 1960.(1)

The new Soviet scheme makes wide use of the techniques which predetermined the success of the decimal classification: the decimal structure of notation for the subdivisions of each class, the synthetic principle, the minutely elaborated tables of common subdivisions. At the same time, the mixed (numerical-letter) notation employing different punctuation marks appears to be rather complicated in comparison with the UDC. Below are the main divisions of the scheme:

А	Marxism-Leninism
Б	Natural sciences in general
В	Physico-mathematical sciences
Г	Chemical sciences
Д	Earth sciences (geodesical, geophysical, geological and geographical sciences)

---

1. Bibliotechno-bibliograficheskaya klassifikatsiya. Tablitsy dlya nauchnykh bibliotek. Vyp. I-25. M., 1960-1969. (Gos. b-ka im. V.I. Lenina). (Bibliothecal-Bibliographical Classification. Schedules for research libraries. Issues 1-25. Moscow, 1960-1969. ⁄Lenin State Library of the USSR⁄).

Е	Biological sciences
Ж/О	Engineering and technological sciences
П	Agriculture and forestry. Agricultural sciences
Р	Health protection and medical sciences
С	Social sciences in general
Т	History. Historical sciences
У	Economics. Economical sciences
Ф	Communist and workers' parties. Socio-political organizations of labour.
Х	Government and law. Juridical sciences.
Ч	Military science.
Ц	Culture. Science. Education.
Ш	Philological sciences. Literature.
Щ	Art. Art criticism.
Э	Religion. Atheism.
Ю	Philosophical sciences. Psychology.
Я	Literature of universal content.

Specific features of hierarchical classifications

Classification is one of the major tools of humanity in the process of cognition, one of the habitual methods which people use to define an object. V.I. Lenin wrote in this connection: "What is meant by giving a "definition"? It means essentially to bring a given concept within a more comprehensive concept".(1) This is one of the strong sides of hierarchical classification as an information retrieval language.

However, it should be recognized that hierarchical classifications, like any other IR languages, have a number of limitations. You will recall that the development of science shows two conflicting tendencies, namely differentiation and integration, or the division of science fields into ever new trends and disciplines, on the one hand, and the interpenetration of the related and even remote fields and disciplines, on the other. The process of science differentiation is more or less adequately recognized in the linear hierarchical classifications. As regards the integration and interpenetration of sciences, however, they cannot be adequately reflected in these classifications. For example, it is an extremely complex task to find in the UDC a place for the scientific trends and topics which emerge at the juncture of chemistry, geology and biology, or mathematics and linguistics.

---

1. V.I. Lenin, Collected works. Vol. 14. Moscow, Foreign Languages Publishing House, 1962, p. 146

This limitation of hierarchical classifications, which strictly delimit the separate sciences in accordance with formal rules of logic, runs counter to the synthesizing trend in the progress of science. The same rules will not permit multi-aspect indexing of documents by means of an hierarchical classification, and information searching for any combination of characteristics. It would be wrong to assert categorically that this constitutes a fault of hierarchical classifications. This is their inherent quality, which renders them highly effective for broad thematic searching when conducted in the conditions of traditional realization in the form of card catalogues. As any other IR language, the hierarchical linear classifications have a restricted range of applications.

Besides, they are undergoing certain modifications which are meant to overcome these limitations. Already the UDC offers some facilities for synthesizing the notation, which enables to a degree to recognize different aspects of a subject classified. Thus, using the relation (colon) sign we can index the multidimensional subject "Bibliography of atomic physics" as 016:539, where 016 stands for Special Subject Bibliographies, and 539, Atomic Physics. The synthetic notation principle had a very significant role to play in the further development of classification.

One of the leading figures in the classification theory, British information scientist D.J. Foskett has this to say in this respect: "There is nothing to be gained by discharging hierarchical classification altogether, provided that we recognize that classification, in the modern sense, can and should mean more than this; that it can cope with the most detailed forms of subject analysis. ... It is evident that a scheme of terms, whether systematic or alphabetical, however well equipped with cross-references, cannot hope to predict all the contexts in which every term may appear, at some time or another. It can, however, provide a set of roles - operating procedures by means of which such contexts may be freshly created out of the scheme as occasion demands.

This means that a classification scheme should no longer set out to provide a "place" for every document, in the sense that a term, or set of terms, will be found in the actual schedules of the scheme for every subject that may be found in a document. In a modern scheme, the art of the classifier in a library is to construct a symbol that is, in effect, a translation of a subject".(1)

Faceted classifications

The first step in this direction was made already by the UDC. The introduction of the common subdivisions and special auxiliaries and the connective symbols, in particular the colon, has appreciably enlarged its capacity for building numbers for complex and multi-aspect subjects. But it is to the eminent Indian library scientist S.R. Ranganathan that we owe a fundamental and consistent approach to the solution of this problem. The Colon Classification which he

---

1. Foskett, D.J. Some fundamental aspects of classification as a tool in informatics. - In: On Theoretical Problems of Informatics. FID 435. Moscow, 1969. p. 65.

developed in 1933 was a further elaboration of the synthetic principle in classification.

S.R. Ranganathan opposed the practice of building minute "enumerative" classification tables in which their compilers sought to provide a separate number for every subject and concept. Instead of a single order to divisions in every main class, he worked out tables based each on some one characteristic or aspect which were later termed "facets". Classification number for any document is built from the symbols used in every table, which are joined by means of the colon. Hence the name - Colon Classification. The classification scheme has not found wide application, but the idea of facet analysis, which forms its most important premise, gave a powerful spur to the development of the library classifications called faceted or analytico-synthetic.

The facet analysis is essentially as follows. At first, a field of science or technology is thoroughly analysed, a faceted classification scheme is built, and a collection of documents in the relevant field is studied. The analysis yields a list of main facets, while the documents provide the essential terms in the subject field, which are grouped by the appropriate facets. Each term in a facet is called a focus. A special significance is attached to the order of the foci within a facet, and the facets within a classification scheme. This order is called the facet formula. S.R. Ranganathan suggested a facet formula that includes five categories:

Personality (class, subclass, subject)
Matter (material)
Energy (Operation, process, action)
Space (place, territory)
Time.

A number of Indian classificationists and a group of British scientists are presently engaged in the further reworking of this formula, increasing the number of the categories it is to contain. The indexing procedure in a faceted classification begins with the formulation of the main subject-matter of a document. It is expressed by a chain of foci which are taken from the facets and arranged in a fixed order. Often instead of the foci their numbers will be used. Such a procedure enables classes to be formed for those documents the subjects of which are expressed by a combination of characteristics viewed from different aspects. To illustrate this rather complex procedure, we shall present an example from the field of medicine (denoted in Ranganathan's system by the letter L):

Facet O	Facet P	Facet H
Organs of human body	Problems of medicine	Care and Treatment
1 Organism as a whole	1 Preliminaries	1 Nursing
2 Digestive system	2 Morphology	2 Etiology
23 Esophagus	3 Physiology	3 Symptom and diagnosis
24 Stomach	4 Disease	4 Pathology
25 Intestine	42 Infection	

3 Circulatory system
4 Respiratory system
45 Lung

421 Tuberculosis

In classifying documents dealing with the "diagnosis of infectious diseases of the intestine", they will be indexed L 25:42:3, and the "pathology of the lung tuberculosis", L 45:421:4.

Compared with the hierarchical classifications of the enumerative type, the faceted classifications offer a number of attractive features: they greatly facilitate the multi-aspect indexing of documents, enabling to bring together in one place all aspects under which a subject or theme is discussed; they are more hospitable to new terms; they provide greater depth of indexing with shorter notation. However even the faceted classifications will not ensure searching by any combination of characteristics, as this would mean that the classified catalogue is to provide places for all possible groupings and regroupings of the foci taken from the different facets. With card catalogues it is an equally complicated thing as in hierarchical classification. And another point to make, the present-day level of development of faceted schemes permits their effective use only in very specialized document collections.

We have thus spoken of the principal types of conventional IR systems: author, subject and classified. Familarity with the rules for bibliographical description and the filing of entries in an alphabetical sequence, the techniques of subject cataloguing and classification open the way to mastering the IR systems, which have evolved in the course of centuries of development and will for a long time to come serve as mass-scale tools of information work. Naturally, the active utilization of these systems requires that a deeper understanding of their languages and the practical experience in their use be acquired. However, if you have formed an idea about the merits of these conventional systems and the limitations in their use which are imposed by their structure it will be easier for you to proceed to the study of the descriptor-type IR systems, which are of fairly recent origin. They will be discussed in the next lecture.

Checkup questions

1. What is the purpose of the basic types of conventional IR systems and what are their specific features?
2. In what context do the problems of bibliographical entry occur and how are they resolved?
3. What properties distinguish the IR languages of subject IR systems?
4. What is the structure of a hierarchical library classification?
5. What are the merits and demerits of the UDC?
6. What are the advantages of the faceted schemes over the hierarchical classifications?

Literature

1. Needham C.D. Organizing Knowledge in Libraries. An Introduction to Classification and Cataloguing. London, Deutsch, 1964. 259 p.
2. Metcalfe J. Subject Classifying and Indexing of Libraries and Literature. New York, Scarecrow, 1959. XII, 347 p.
3. Guide to the Universal Decimal Classification (UDC). B.S. 1000C:1963. (FID No. 345). London, British Standards Institution, 1963. 128 p.
4. Foskett D.J. Some Fundamental Aspects of Classification as a Tool in Informatics. - In: On Theoretical Problems of Informatics. FID 435. Moscow, 1969, p. 64-79.

7. DESCRIPTOR-TYPE INFORMATION RETRIEVAL SYSTEMS

We have made it clear in the preceding lecture that the conventional information retrieval systems, both subject and those based on hierarchical and faceted classifications, along with definite advantages suffer from a certain limitation : they do not provide for document searching by any combination of characteristics that has not been established beforehand. The conventional IR systems, in particular the alphabetical subject systems and hierarchical classifications, become too cumbersome when confronted with the necessity of providing for multi-aspect searching, i.e. searching for documents by multiple characteristics belonging to the different aspects in which a subject or phenomenon is viewed. This limitation becomes even more pronounced when a search for information is carried out in a file of documents that have many such aspects, and the aspects do not fit into a natural hierarchy of generic relations.

An instance of the limitation of conventional IR systems

To remind you of this feature of the conventional IR systems, we shall use as an illustration a homely and deliberately simplified example where these systems appear in a schematic form. Let us assume that we conduct an information search in a file of motion pictures according to four aspects containing two characteristics each. The motion pictures are normally divided, according to content, into documentary and art films; according to format, into normal and wide-screen; according to colour, into black-and-white and colour; according to length, into short and full-length.

In a subject system the name of each of these eight characteristics will form a heading and all of them will be arranged alphabetically:

Art films
Black-and white ...
Colour ...
Documentary ...
Full-length ...
Normal ...
Short ...
Wide-screen ...

The divisions under each of these headings will list the address codes (numbers) or titles of those motion pictures which possess the relevant characteristics. You remember that this system is easy to use, practically does not require any preliminary study of the principles of its build-up on the part of the user, and will quite readily accept new characteristics. For example, if we were to index educational or popular science films, stereoscopic or cinerama films, we had only to insert into the alphabetic listing the subject headings for these characteristics. Then in answer to a request for a film possessing any of these features the system will promptly respond with the needed information.

The situation is radically changed, however, when you want to choose a film by a combination of several characteristics. The system is unable to provide an answer to a multi-aspect query such as this: "what colour wide-screen art films or films with alternating colour and black-and -white images are available?" To supply an answer, all the relevant subject headings would have to be looked up and the matching film titles identified. The subject system is not designed for a task like this. We know that a way of overcoming this limitation is to form compound subject headings which combine several characteristics to provide for the would-be requests. However, it is not always possible to foresee all possible requests, and apart from this, the system then becomes too complex and cumbersome.

Now let us try an hierarchical classification. Taking the first letters of the listed characteristics as the codes of the respective subject headings, and observing the formal rules of construction of hierarchical classification, we shall obtain the classificatory tree presented in Fig. 26. This example obviously demonstrates that the hierarchy of generic relations will not at all necessarily be a natural one. Here the choice of the aspects to serve as the characteristic of division at the upper levels of hierarchy depends on the nature of the potential information requests. But they are hard to predict, for they depend both on the purpose of the show, on the film-goers' makeup, on the time they are ready to spend, and on the projecting equipment available.

With the classification chosen in Fig. 26 the system is unable to provide an answer to a multi-aspect request on the availability of short or full-length pictures (eight subject headings will have to be scanned) or the black-and-white and colour pictures (four headings will have to be scanned). It is also clear that the necessity of adherence to the formal rules of logic in the construction of a hierarchical classification contributes to the formation headings that are very rarely used. Thus, to stick to the above example, the headings ANBS, ANCS, AWBS and AWCS will not contain very many entries, since short art films are produced but infrequently.

To enable the system to give answers to any request using any combination of characteristics, it would be necessary to employ successively each aspect of each characteristic at each level in the hierarchy. This is impracticable because of the very great number of the subject headings that would be formed. Another important point to make is that in order to introduce new characteristics, e.g. for the indexing of educational and popular science films or stereoscopic and Cinerama films the system would have to be considerably reworked to accommodate new classes and subclasses with all the subordinate headings.

Motion pictures where the colour and black-and-white images are alternating would be entered under several subclasses at once. To enable the system to retrieve only these films, new subclasses uniting both characteristics would have to be formed within it, thereby further increasing its complexity. The system would be made even more complicated if it were necessary to incorporate into a hierarchical classification news aspects of treatment of films, for instance, according to country of origin (domestic, foreign), or according to the age group of the film-goers for which they are meant (for the children, for the adults), and so on. It would result in new levels of hierarchy and many new subject headings.

The faceted classifications are able to make good some of these inadequacies. Since no attempt is made in compiling them to list as many combinations of characteristics as is possible, but instead a sort of building blocks are offered in the shape of foci (characteristics), joined in facets (aspects), the indexer can omit the practically inessential characteristics without running the risk of disturbing the hierarchy, as well as form new subject headings of a specific interest. In our example, for the indexing of short documentary newsreels a generalizing subject heading D:S could be set up, and for the art films we could make do with the subject headings A:N:B; A:N:C; A:W:B; A:W:C. If necessary, to these can be added, leaving the system intact, the subject headings for the motion pictures with alternating colour and black-and-white images: A:N:BC; A:W:BC, as well as subject headings for new characteristics (foci) - educational films, Cinerama films, etc.

It would not be too difficult to insert into the faceted classification also new aspects of treatment of pictures, e.g. according to their place of origin, the film-goers age group, etc. However, a faceted classification will not ensure a search for any combination of characteristics either, because for this it would have to comprise no fewer subject headings than an hierarchical classification. In our example the system fails to answer a query on all short colour films or black-and-white wide-screen films.

Therefore, the conventional IR systems, which have for centuries been evolving to provide answers to broad thematic and single-aspect requests, have proved to be poorly equipped for specific and multi-aspects searching, as well as searching for any combination of characteristics that has not been previously established. The most extensively used conventional IR systems, those based on hierarchical classifications, have much difficulty in housing the ever increasing number of multidisciplinary problems which often include aspects having no clear-cut generic relations.

These circumstances furthered the emergence, some twenty years ago, of a new method of information retrieval which was to be called coordinate indexing. This method forms the basis of the descriptor-type IR systems, discussed in this lecture.

Set-theoretic terminology

In setting forth the essentials of coordinate indexing, it is convenient to use the elementary terms of the set theory. For this reason it may be useful to introduce here some terms pertaining to the set theory and the logical operations on sets. A set is a collection of objects. Letters of the alphabet papers in periodicals, books on shelves, the numbers 1,2,3,4,5, - all are examples of sets. A set may contain only one element or be empty (contain no elements whatsoever). Sets may also be infinite, e.g. the infinite number of points on a circle. A set is generally denoted by writing its elements within braces e.g. $\{1,2,3,4,5\}$. The statement "A is a subset of B" (i.e. every element of A is an element of B) is written in a contracted form as $A < B$ or $B > A$ (the inclusion relation). An empty set is denoted by the zero; a set which comprises all the objects in a given field is called universal.

The main operations on sets are the logical sum, logical product, logical difference, and logical complement. The union (sum) A B of two sets A and B is a set each element of which is an element of A or of B or of both. The union of the set $\{1,2,3,\}$ and the set $\{2,3,4,5\}$ is the set $\{1,2,3,4,5\}$. The intersection (product) $A \cap B$ of two sets A and B is a set each element of which is simultaneously an element of A and of B. The intersection of the sets $\{1,2,3\}$ and $\{2,3,4,5\}$ is the set $\{2,3\}$. The difference, A - B, between two sets A and B is a set of all elements of A that are not elements of B. The difference of the sets $\{1,2,3\}$ and $\{3,4,5\}$ is the set $\{1,2\}$. The complement, A/B, of a set A into a set B is a set of all elements of B that are not elements of A. The complement of the set $\{1,2,3\}$ into the set $\{2,3,4,5\}$ is the set $\{4,5\}$. The complement, $\bar{A}$ or A', of a set A in the universal set is a set of all elements of the universal set that are not elements of A. The graphic representations of the relations between sets are given in Fig. 27.

Coordinate indexing

The coordinate indexing method is based on the assumption that the salient contents (subject) of a document and information request can with sufficient accuracy and completeness be expressed by an appropriate list of so-called keywords which are explicitely or implicitly contained in the text being indexed. By keywords we mean the most essential words for expressing the main meaning of a word or phrase, which have a nominative function. Most of the nouns, adjectives, verbs, adverbs, numerals and pronouns can be used as keywords. Prepositions, conjunctions, connectives, particles and other function words cannot be keywords. To put it differently coordinate indexing is a method of expressing the primary subject-matter (subject) of a document or information request by a certain totality of keywords. In "pure" coordinate indexing, the keywords in the search patterns are not related to each other and function independently. In contrast to the search patterns of documents the search request formulations are presented as logical sums, products or complements of the classes designated by the corresponding keywords. In order to locate a document matching an information request it is necessary to certain logical operations on the classes designated by the keywords in the search patterns of the documents.

In the simplest case, when a search request is formulated as the logical product of a certain set of keywords, the document is held to be a match if its search pattern contains all the keywords of the search request formulation. In a sense, this is equivalent to the operation of logical product of classes designated by these words of the search pattern which coincide with the keywords constituting the search request formulation.

To illustrate this point we shall again pick up the example instanced in the beginning of this lecture. Suppose that our collection of films is indexed by keywords naming the characteristics that have been chosen. To cull out from this clllection the short films with alternating black-and-white and colour images (a question which would present great difficulties to any conventional IR system) it will be sufficient to state the request as the logical product of the three keywords describing these characteristics: $S \cap B \cap C$. Another, more abstract example was given in Fig. 16, at the beginning of the fifth lecture. There we conducted a search by means of keywords lettered A to H. We wanted all documents containing characteristics C and D or C and F. Using the terminology that we have introduced we can now say that the request formulation was composed as a logical sum of two products of keywords $(C \cap D) \cup (C \cap F)$.

Externally, the coordinate indexing method appears to ressemble an IR language of a subject system. In either case we are dealing with classes designated by appropriate natural language words and phrases, that is, by keywords and subject headings. Each keyword, e.g. "electric", covers a class of subjects and concepts the names of which include this word. The same is true of the formulation of subject headings. But the similarity ends here.

In coordinate indexing, in order to locate a document whose search pattern is composed of a certain totality of keywords, it is necessary to perform the logical operations of sum product or complement on the classes designated by these keywords. This provides an opportunity of conducting a multi-aspect search for any previously unspecified combination of keywords. In the subject systems, on the other hand, as well as in the hierarchical and faceted classifications, each subject heading or classification number appear independently, listing the address codes (numbers) of all relevant subject materials. In conventional IR systems simultaneous searching for several subject headings or class numbers is a difficult matter indeed, and sometimes a downright impossibility, as they are not designed for such use. For example, a search for documents using any combination of five mutually exclusive characteristics will require, in a system based on coordinate indexing, a search request formulation as a logical product of five keywords. In any conventional IR system the same task will involve the scanning of the contents of as many subject headings as there are permutations of five headings or classification numbers, i.e. $5! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 120$. This constitutes one of the most important advantages of coordinate indexing over the languages of the conventional systems.

However the "pure" coordinate indexing procedure we described is inadequate, as it does not ensure the acceptable levels of recall and precision. Here are some of the main reasons for this. The natural language words from which keywords for coordinate indexing are chosen have the property of synonymy. One and the same thing or idea may be denoted by different words. Such words which are mutually replaceable in similar contexts without giving the feeling of a change in meaning are called synonyms. Examples of synomyms are such words as "test" - "trial", linguistics" - "science of language", "common salt" - "sodium chloride". It can be seen that if the same subject is designated by different words in the search pattern of a document and in a search request formulation the document, though relevant to the request, will not be retrieved.

Another source of the inadequacy of "pure" coordinate indexing is the multiple meaning of natural language words. It manifests itself in polysemy, which is a transfer of meaning from one subject to another having common qualities or characteristics, and in homonymy, which refers to the accidental coincidence of different words that have nothing in common either in respect of origin or meaning. Examples of polysemantic words are "star" - a celestial body or a geometrical figure, "bureau" - a kind of table and a collegiate body, examples of homonyms are "mail" - the system of communication and an armour; "ray" - a beam of light and a sea-fish. Evidently, the multiple meaning of keywords may bring about the retrieval of non-relevant documents which means lower precision values and increased "noise".

One more source of insufficient effectiveness of information retrieval by coordinate indexing may lie in the unspecified relationships between keywords which results in the non-retrieval of relevant documents, that is in poor recall. Let us assume that a search request is for documents discussing the properties of liquids, and the search request formulation includes the keyword "liquid". If it is in no way indicated that water is a liquid, the documents dealing with the properties of water will not be retrieved. The most important relations between keywords are the generic relations, but certain significance is attached to other relations as well, particularly to the associative relations, i.e. the relations between keywords by association: by contiguity ("chair" - "table", "fall" - "autumn"), by similarity ("clock" - "scales" - "thermometer", "pond" - "lake" - "sea"), by contrast ("black" - "white", "light" - "heavy"), etc. The sum total of the conceptual relations between keywords, based on the relations between the ideas, things or phenomena they designate, are called paradigmatic relations.

Descriptor information retrieval language

It follows from the examples discussed above that in order to effect a considerable increase in the precision and recall of information retrieval based on coordinate indexing it is necessary at least to establish control over the vocabulary of the IR language used. Such a control should consist in the elimination of synonymy and in the building of special distionaries, tables or charts which would graphically demonstrate the most essential paradigmatic relations between the keywords. With full vocabulary control, the coordinate indexing of search requests and documents will make use of only those words which are contained in a certain normalized list in which their synonymy and polysemy has been eliminated and the paradigmatic relations shown. Such keywords have been given the name of descriptors.

Descriptors are normalized keywords designed for coordinate indexing of documents and information requests which have been selected according to certain rules from the main vocabulary of a particular natural language and which have been, by some artificial means (cross-references and notes) freed from synonymy, polysemy and homonymy. A descriptor language is a specialized IR language the vocabulary of which is composed of descriptors, and a grammar, in the simplest case at least of a method of building search patterns and search request formulations by correlating relevant descriptors.

It is impossible to establish now who was the originator of the coordinate indexing idea. The British information scientist B.Vickery has noted that coordinate indexing with inverted file structure was apparently known in Sumer three thousand years ago. Clay tablets were found, each assigned to a symptom of a disease; the names of all diseases characterized by one or another symptom were listed on the respective tablets. Obviously, if according to the symptoms observable in a patient the corresponding tablets are selected and the name of the disease inscribed on each of them is found out, it will in all probability be the disease with which the sick man in question is afflicted. These Sumerian tablets are the archetype of modern diagnostic machines.(1)

The coordinate indexing principle forms the foundation of the so-called superimposable cards, which will be considered in the next lecture. In 1915, they were used by ornitologist H. Taylor, in the 20's they found application in the identification of minerals and in personnel selection, and in 1939, W. Batten, of the Imperial Chemical Industries (Britain), used them for patent searching. However, the credit for the development of the information retrieval systems which formed the basis for the development of modern descriptor languages, goes to the American scientists C.N. Mooers and M. Taube.

C.N. Mooers designed in 1947 a mechanized document searching device which he named "Zatocoding system". It was he who coined many of the terms which are now in wide currency among information scientists, including "information retrieval", "information retrieval system", "retrieval language", "descriptor", "descriptor dictionary". In one of his earlier works C.N. Mooers has thus stated the indexing method in the system that he developed: "Each document or unit of information is characterized by a set of descriptors taken from the vocabulary of descriptors. Each descriptor of the set applies to, or is true in some way, of the information content of the unit of information. The descriptors operate independently in this type of characterization. The fact that there are several descriptors in the set may mean that they formed some interacting combination in the original document, or it could just as well mean that they relate to independent ideas scattered through the document. Using descriptors in this fashion drops almost all relationship between the ideas represented by the descriptors".(2)

M. Taube had a great contribution to make to the theoretical substantiation, development and popularization of the ideas of coordinate indexing. He defined coordinate indexing as "a method of analyzing items of information so that retrieval is performed by the logical operations of the product, sum and complement on the codes in the store".(3) In 1951 M. Taube developed the so-called Uniterm system which has found extensive application since then. The main difference between the Uniterm and Zatocoding systems is that the former deals with words expressing concepts while the latter deals with concepts expressed by words.

---

1. International Study Conference on Classification for Information Retrieval. Dorking, 1957. Proceedings. Lnd., Aslib, 1957, p. 106.

2. Mooers C.N. Zatocoding Applied to Mechanical Organization of Knowledge. - "American Documentation", 1951, v. 2, N 1, p. 26.

3. Taube M., Wooster H. (eds.). Information Storage and Retrieval. Theory, Systems, and Devices. New York, Columbia Univ. Press, 1958, p. 8.

Uniterm system

M. Taube's Uniterm is a keyword (as a rule, a simple one) which may have an appropriate cross-reference or scope-note which helps to eliminate its multiple meaning. By contrast with C. Mooers' descriptors, the Uniterms have no references which would specify the paradigmatic relations between them. Uniterms may be keywords expressing single ideas, as well as proper, geographical and trade names. All Uniterms have an equal hierarchical rank: none of them occupies a conceptually superior position with respect to any other Uniterm (as in hierarchical classifications) and none is used in a preestablished combination with any other Uniterm (as in subject headings).

This method of building an IR language vocabulary contributes to a great reduction in its size. The list of subject headings in a Library of Congress subject catalogue contained some 50,000 entries. Conversion to Uniterms resulted in a list containing a mere 3,000 words. The important feature of the Uniterm system is that the vocabulary is built in the process of using it and not developed beforehand, as in the Zatocoding system. During the initial period of the system operation the number of Uniterms will grow steadily. Gradually, synonyms and polysemantic words will be detected and marked on uniterm-cards. With the growth of the document collection the rate of growth of the vocabulary will tend to slow down eventually coming almost to a standstill.

Using as an example this system - one of the first and simplest systems based on the coordinate indexing method - it will be easy to get acquainted with the practical techniques of input and searching in a descriptor-type IR system. For the physical implementation of his system M. Taube developed a special Uniterm card having either the usual catalogue card size (75mm x 125mm) or larger in size (203mm x 125 mm), on which a grid formed by one horizontal and ten vertical lines is printed. The horizontal line is used for writing a Uniterm. Uniterm cards are envisioned for each Uniterm used in the system and filed alphabetically, with due account taken of those Uniterms which consist of more than one word.

The vertical columns are for posting the codes (numbers) of documents the search patterns of which include the Uniterm indicated in the horizontal line of the card. The peculiarity of the system is the "terminal digit" order of postings on the card. This means, for example, that number 127 is posted in the 7 column, number 239, in the 9 column, and number 270, in the 0 column, etc. Since the documents are numbered as they are entered into the system, they will be posted in the columns in the ascending order of their numbers. Such a system of recording appreciably facilitates the visual scanning and locating of document numbers which are common to all the cards being matched. This procedure in this case amounts to the operation of logical product of the Uniterms contained in a search request formulation. The method of work with Uniterm cards is demonstrated in Fig. 28.

Let us assume that we are to enter in the system a paper dealing with methods of "corrosion protection of gas turbine blades". For the coordinate indexing of this paper we need the Uniterms TURBINES, GAS BLADES, CORROSION, and PROTECTION which jointly form its search pattern. The paper is assigned its serial address code 2005. Then the cards assigned to these Uniterms are removed from the file. In doing it the Uniterm PROTECTION or its synonyms are found to be not in the system, and a card is made for this particular feature. Number 2005 is posted in the 5 column in each of these Uniterm cards, following which they are replaced in the file. Similar procedures are performed for each incoming document.

An information search in a file of Uniterms is conducted in the following way. Suppose that a request is put to the system for papers dealing with "gas turbine blades". The search request formulation is made to contain the Uniterms GAS, TURBINES, BLADES. The cards for these features are extracted from the Uniterm file and the numbers which are common to all of these cards are located. It is advisable to have as the basis for comparison the card which contains the least number of postings (in this case, the Uniterm card for BLADES). Matching should likewise begin with the columns with the least number of postings. The columns that are empty in one of the cards need not be matched at all. Where the base card has columns with one posting it is not expedient to match the postings which have been made before this single posting (i.e. smaller numbers). The time it takes to locate all common postings is usually negligible. Thus, in our example it can be seen at once that the three topmost Uniterm cards have only three matching numbers: 526, 1027, and 2005, which are the address codes of the documents sought. If the request is broadened to cover documents on gas turbines (two cards at the top), the output will increase to include five documents: 195, 294, 526, 1027, 2005. If, on the other hand, the request is made more specific to read "gas turbine blades corrosion" (four uppermost Uniterm cards) only two relevant documents will be retrieved, namely 1027 and 2005.

Thesaurus and its construction

In the Uniterm system, the control over the retrieval language vocabulary, as we have seen, is limited to the elimination of synonymy and polysemy. However, full lexical control, which can alone ensure high precision and recall values, especially with thematic searches and searches for documents which match the request but partially, necessitates the recognition of paradigmatic relations between the indexing terms. For this purpose special normative references dictionaries called thesauri are compiled.

An information retrieval thesaurus is a reference dictionary designed to help the information user to state his information needs in terms of the descriptor language and to provide for redundant indexing of documents and information requests by these terms. It must contain all the descriptors used by the language of the given system displaying in a graphic form their conceptual relationships, as well as the keywords within this system which are considered to be synonymous with these descriptors. Therefore, the thesaurus will help to eliminate synonymy and polysemy of keywords and the lack of explicit relations between them, which give rise to the defects of "pure" coordinate indexing that have been discussed before.

Synonymy of keywords is eliminated in this way. In constructing a thesaurus a list of keywords used for the coordinate indexing of documents is first compiled. Then groups of words which can be taken to be synonyms are brought together. From each of these groups a word or phrase is taken to represent the whole group which becomes a descriptor. The rest of the words in the group are held to be synonyms of the descriptor and linked to it by "see" and "use" references. Each descriptor is joined with all its synonyms by backward "includes" references, e.g.

ABSTRACT includes	speculative theoretical
speculative see	ABSTRACT
theoretical see	ABSTRACT

Polysemy of keywords is also taken care of in building up a dictionary of descriptors. For this each multiple-meaning keyword is affixed alphabetic of numerical symbols and words qualifying its meaning. Another method is to replace polysemantic keywords by descriptors composed of one-valued phrases, e.g.:

atlas (geographical) see geographical atlas
atlas (vertebra) see cervical atlas
filter - 1 (gas)
filter - 2 (optical)
filter - 3 (chemical)
filter - 4 (electrical)

The conceptual (paradigmadic) relations between descriptors are also displayed in the thesaurus. According to our definition, paradigmadic relations are such relations between descriptors which are based on the existence of objective connections between the ideas, objects or phenomena they denote, In this case synonymy is left out of account, because in a descriptor language it is eliminated and does not exist within one system. The following paradigmadic relations can be cited as the most important ones:

species - genus (genus - species);
collateral subordination;
similarity (functional)
cause - effect (effect - cause);
part - whole (whole - part).

Of particular importance are the generic relations and the collateral subordination relations realizable through them. It is these relation types that form the groundwork of hierarchical classifications. Almost every concept can be at once generic (i.e. it can reflect the salient features of a class of objects that includes other classes of objects which are species of this genus) and specific. The notion of "rectangular", for example, is generic to the notion "square" and specific to the notion "parallelogram". Exceptions are constituted only by the broadest concepts, or categories (e.g "matter", "space", "time") which have no generic concepts, and the most narrow, single notions (e.g. "UDC") which have no specific concepts. The concepts which are equally subordinated to one generic concept are said to be collaterally subordinated, e.g. the notions of "phonetics", "lexicology" and "grammar" are collaterally subordinate with respect to the generic notion of "linguistics". It should be noted that in contrast to hierarchical classifications, descriptor languages take into consideration generic relations irrespective of the level of hierarchy to which they belong.

The thesauruses indicate generic relations independently of the other paradigmadic relations which may outwardly resemble them. This applies to the functional similarity relations (clock - scales - thermometer), causal relations tiredness - sleep), and particularly to the relations of the "part - whole" kind. It can be said of the latter type of relations that they differ from the generic relations in that they represent relations between objects and not concepts. "Genus" and "species" are abstractions, while "whole" and "part" are concrete things.

Thus an information retrieval thesaurus generally consists of three parts:

1. The vocabulary part is a common alphabetic list of descriptors and keywords viewed in this system as synonyms of these descriptors. Descriptors are commonly made to appear more prominent in the listing, e.g. by printing them in capitals, and are linked by cross-references to all their synonyms. Polysemy and synonymy of keywords are eliminated using the method described above.

2. The "semantic map" of the retrieval language vocabulary is a network of conceptual classes in which all the descriptors of a given language are grouped. This part of the thesaurus provides a graphic demonstration of the essential paradigmadic relations between descriptors, at least of their generic relationships. These relations are expressible in one of the two ways: either by the combinations of the alphabetically ordered thematic classes (fields) containing a multiplicity of thematic groups of descriptors also ordered alphabetically, or by charts in which the basic paradigmadic relations are indicated by arrows.

3. The rules of conversion of keywords and key phrases of the natural language into a descriptor-type IR language determine the procedure of substitution of descriptors for these keywords and phrases. The rules define the conversion of the names of institutions, chemical compounds, biological species, and other similar categories of terms, as well as the rules of vocabulary editing of search patterns and search request formulations, e.g. for their complementation with descriptors connected with the main descriptors by generic and other paradigmadic relations.

To illustrate the thesaurus structure two examples will be instanced. The first published thesaurus to use the graphic method of displaying paradigmadic relations between descriptors was the "Euratom-Thesaurus" (first edition) for nuclear physics and engineering.(1) It contains 4,470 descriptors, including 1,836 names of inorganic chemical compounds and 1,404 names of isotopes. The 42 subject classes, numbered 00 to 94, with lacunes, are in the form of charts in

---

1. Euratom-Thesaurus, Keywords used within Euratom's Nuclear Energy Documentation Project. EUR 500.e (1st ed.). Brussels, 1964. 80 p. (European Atomic Energy Community).

which the arrows are directed from the generic to specific descriptors. The collaterally subordinate descriptors are marked by bilateral arrows. These classes cannot be considered as classifications; they broadly accord with those subject fields that are of interest to Euratom. The vocabulary part provides a common list of descriptors with references to the relevant subject classes and a common list of their synonyms which are associated with the corresponding descriptors and their subject classes. The generic relations are thus completely ignored in the vocabulary part. Fig. 29 shows the subject class 71 Mathematics from the Euratom-Thesaurus, first edition. The arrows which go beyond the limits of a subject class join its descriptors with the other subject class descriptors and their numbers. To the right is an alphabetic listing of all descriptors which belong to this particular class, and at the bottom there are numbers of all subject classes used in the thesaurus.

Another example to cite is the "Thesaurus of Engineering Terms" which was brought out in 1964 by the Engineers Joint Council (USA).(1) By contrast with the previous thesaurus it indicates generic and other paradigmadic relations by means of a system of references and scope notes. The thesaurus contains 10,515 words of which 7,750 are descriptors. The listing of the descriptors and their synonyms in the vocabulary part is alphabetical. The synonyms are joined to the descriptors which are meant to replace them by the "use" references. In the "semantic map" which has the form of an alphabetic listing of the basic (heading) descriptors each entry includes the heading descriptor, its synonyms, the specific descriptors, the generic descriptors, the descriptors connected with it by other paradigmadic relations. To designate them the following notes are employed:

UF (used for) - for synonyms;
NT (narrower term) - for specific descriptors;
BT (broader term) - for generic descriptors;
RT (related term) - for other descriptors.

The ampersand (&) is used to mark those descriptors which can be used in place of other descriptors expressing more narrow concepts and marked by the symbol #

The second major thesaurus published by the Engineers Joint Council, Thesaurus of Engineering and Scientific Terms 1st ed. (New York, 1967), uses some other symbols. The dagger (+) in front of a term signifies that two or more descriptors are to be used in coordination for that term. The dash (-) symbol in front of a descriptor indicates that the descriptor has narrower terms (not shown) and that the main entry should be consulted to determine these. Below is a simple dictionary entry from the Thesaurus of Engineering and Scientific Terms:

---

1. Thesaurus of Engineering Terms: a list of engineering terms and their relationship for use in vocabulary control, in indexing and retrieving engineering information. 1st ed. New York, Engineers Joint Council, 1964.

Scientist

UF Scientific personnel
BT Personnel
Professional personnel
NT chemists
Physicists
RT - Engineers.

A few brief words will be in order on the term "thesaurus" (from the Greek "thesauros", literally meaning a treasure, treasury or storehouse). The word seems to have been first used in its present meaning by Florentine Brunetto Latini (1220-1294) in his encyclopaedia "Li Livres dou Trésor". In the 16th century the word appeared in the names of the Latin and Greek lexicons "Dictionarium, seu Lingue Latinae Thesaurus" (1532) and "Thesaurus Linguae Graecae" (1572) which were published by the well-known French philologists and publishers the Estiennes. In contemporary usage the term refers to the dictionaries of concepts, or ideological dictionaries, which are in effect inverted explanatory dictionaries. Of these, the greatest popularity is enjoyed by the "Thesaurus of English Words and Phrases" which was compiled by P.M. Roget in 1852 and has passed through at least 90 printings since. Similar dictionaries, joining all words in thematic groups or subject classes, are to be found in the French, German, Spanish and other languages as well.

One of the first to recognize the need for this type of dictionary in information retrieval work was the American bibliographer C.L. Bernier who wrote back in 1957: "A limited thesaurus would seem to be another effective way of bringing the relevant terms to the attention of the searcher if the vocabulary proves too large to be read completely each time for selection".(1)

Grammatical means of descriptor IR languages

In using descriptor languages, the search pattern of every document and search request formulation is stated in the form of an unordered set of descriptors. However, in addition to the paradigmadic relations between them, which in some way or other are recognized in the indexing and in searching assisted by a thesaurus, there exist different relationships between descriptors that derive from the document contexts. Such relations between the descriptors are said to be syntagmatic. If these relations are ignored too, the descriptors belonging to a document search pattern may form spurious combinations which result in the retrieval of non-relevant documents, increased "noise" and hence lower precision.

The following examples will illustrate this point. Suppose that our system includes a document on the "production of sulfuric acid and catalyst purification". Its search pattern will contain the following descriptors: PRODUCTION, SULFURIC ACID, CATALYST and PURIFICATION. These descriptors may, during a search, form

---

1. Bernier C.L. Correlative Indexes. II: Correlative Trope Indexes. - "American Documentation", 1957, v. 8, N I, p. 48.

false combinations which will result in the retrieval of this document in answer to a question on the "purification of sulfuric acid" and on the "production of catalyst", although the document does not deal with anything of the kind.(1) Another example. A document deals with the "coating of copper tubes with lead". It is indexed by the following descriptors: LEAD, COATING, COPPER, TUBES. If a request for documents on lead tubes is entered into the system the request formulation will be indexed by the descriptors LEAD and TUBES, and the document will be retrieved, although it is not relevant to the query.(2)

The principal grammatical means which are used in the descriptor languages to reduce "noise" are the links and role indicators, or roles. Roles are special symbols which are attached to a descriptor and reduce the scope of the concept which it stands for. This is achieved by indicating the logical role which the given descriptor plays in the particular context. In the first of the cited examples in order to prevent spurious combinations it will be enough to affix to the descriptor SULFURIC ACID in the search pattern the role indicator "B" showing that it is a product of chemical or industrial reaction, and to the descriptor CATALYST, the link "C" indicating an undesirable component (waste, impurity, admixture, spoilage). The document search pattern will then be indexed: PRODUCTION, SULFURIC ACID-B, CATALYST-C, PURIFICATION, and the request formulation will read: CATALYST-B, PRODUCTION, SULFURIC ACID-C, PURIFICATION. The document will then not be retrieved in response to the queries asking for documents on catalyst production and sulfuric acid purification.

Links are also special symbols attached to the descriptors in the search patterns of documents (or their address codes) and designed for the conceptual grouping of these descriptors. In the second of our examples, the search pattern of the document on the coating of copper tubes with lead would assume this form by the addition of links; LEAD R_1, COATING R_1, COPPER R_2, TUBES R_2. The link R_1 would join together the descriptors LEAD and COATING, and the link B_2, the descriptors COPPER and TUBES. If the condition is stipulated that a document is to be produced on request only when the matching descriptors in its search pattern and in the search request formulation have the same links ascribed to them, this document will not be retrieved in answer to a query on lead tubes. But then it would not be retrieved either in response to a request for documents dealing with tube coatings generally. This would mean an information loss. The search pattern can be made more complex by the addition of several links: LEAD R_1, COATING R_1R_3, COPPER R_2R_3, TUBES R_2R_3. If now the condition states at least one common link in the matching descriptors, this document will be retrieved in response to questions on tube coating, lead coating, and coating of copper, but will not be retrieved by a false correlation in response to a request for lead tubes.

Experience indicates that the best results are produced by the joint utilization of links and roles. They reduce noise by 10-15%, but considerably

---

1. Holm B.E. Information Retrieval - a Solution.- Chemical Engineering Progress", 1961, v. 57, N 8, p. 74-75.

2. Taube M. Notes on the Use of Roles and Links in Coordinate Indexing.- "American Documentation", 1961, v. 12, N 2, p. 98-100.

increase the cost and time of indexing. In relatively small document collections (up to 30,000 items) the increased precision of retrieval achieved through the use of grammatical means will not balance the increased time and cost of indexing and searching. This means that under certain circumstances the employment of descriptor languages without grammar is more advantageous.

Further development of grammatical means of the descriptor languages has resulted in specialized IR languages in which the semantic relations between descriptors are recognized to a greater extent. These relations and the descriptors themselves are designated by complex codes and the search pattern of a document has the form of an encoded "telegraphic abstract". In one such language the notion of "thermometer" is treated as a "machine or device (MACH.) which effects (U) the measurement (MUSR.) and is affected (W) by heat (RWHT.)" and is designated by the descriptor MACH.MUSR.RWHT. Among the most well-known of such languages are the WRU Semantic Code, SYNTOL, the language of RX-codes, and some others. They are in the experimental stage as yet and designed primarily for automating information retrieval with the aid of computers. Familiarization with these languages is beyond the scope of our introductory course.

We have thus considered some of the fundamental problems relating to the descriptor-type IR systems; the reasons for their appearance and their advantages over the conventional systems; the principles of coordinate indexing and its inherent weaknesses; the basic notions of descriptor languages; the operation of some of the simpler systems based on these languages; the building up of the thesauruses and their range of applications, and finally, the grammatical means of descriptor IR languages. To round up the discussion of information retrieval problems we have only the technical means of realization of IR systems to consider. This will be the subject of the lecture that follows.

Checkup questions

1. How do the inherent limitations of conventional information retrieval systems manifest themselves?
2. What is coordinate indexing and what are its advantages over the IR languages of the conventional systems?
3. Give the definitions of a descriptor language and its principal concepts.
4. How is a Uniterm system designed and how does it work?
5. What is a thesaurus, its structure and purpose?
6. What is the role of the grammatical means of a descriptor language?

Literature

1. Taube M. and Wooster H (eds.) Information Storage and Retrieval. Theory, Systems, and Devices. New York, Columbia Univ. Press, 1958.
2. Costello J.C. Training Manual and Workbook for use in Abstracting and Coordinate Indexing. Training Course. Columbus, Battelle Memorial Institute, 1964. IX, 117, 11 pp.
3. Vickery B.C. On Retrieval System Theory. 2nd ed. London, Butterworths, 1965, XII, 191 p.
4. Howerton P.W. (ed.). Information Handling: First Principles. Washington, Spartan Books, 1963.

8. TECHNICAL MEANS OF INFORMATION RETRIEVAL

While discussing information retrieval systems in general terms in the last three lectures we hardly mentioned the technical means of their implementation. The only exceptions were library systems and Uniterms, based on such elementary facilities as bookshelves and catalogue cards. As you may remember, we did this on purpose, being aware that diverse technical means can be used to implement one and the same system and that it is therefore convenient to speak of all those means in one place.

Their turn has come now, but we must first discuss a number of general aspects of mechanization and automation of information retrieval.

Mechanization and automation of information retrieval

Technology is penetrating all spheres of human activity with rapid strides, but the system of scientific communication no doubt still remains in this respect at the level of centuries long past. It was not by accident, therefore, that the first paper to acknowledge the full-weight status in science of scientific information problems urged the development of special information machines. The paper was by an American scientist, engineer and public figure V. Bush, who wrote: "Professionally our methods of transmitting and reviewing the results of research are generations old and by now are totally inadequate for their purpose."(1)

Definite successes have been scored in that respect in the quarter of a century that has passed since. However, we are still far from a solution to the problem of mechanized and automated information retrieval. Year-long futile attempts at designing a "big information machine" finally brought in a realization that this end would be better reached with standard punch card and electronic digital computers used in combination with microfilming and other copying facilities. Though, an aftermath of this"children's disease" was still manifest in an overestimation of the role to be played by machines in solving information retrieval problems. This overestimation survives even today despite the new clear awareness that information retrieval problems can be solved only through a deep-going study of thought processes.

That reminds of the vivid words of French mathematician L. Brillouin who wrote: "Too many laymen picture science as a sheer accumulation of facts, and take a scientist for a living encyclopaedia. They think it possible to solve every problem by means of giant machines which might register all human knowledge in their enormous, magnetic memories. Do you need some information? Push a button and the machine will answer! This is ridiculous, and stems from the poorest

---

1. Bush, V. As We May Think.- "Atlantic Monthly", 1945, v. 175, No. 1, p.101; reprinted in: Readings in Information Retrieval. New York-London, 1964, p.20.

science fiction. More important, it proves that the role of the scientist is too often completely misunderstood".(1)

This, of course, does not imply that mechanized or automated retrieval is unnecessary. But complex facilities should only be introduced into information processes provided they bring tangible economic gains by reducing the operational costs of the retrieval system, appreciably speed up retrieval, or permit some retrieval functions that are in principle unfeasible in manual systems.

It should be kept in mind that machines are not automation's Problem No.1, and that it is not with them that automation must start. Machine may be charged only with those operations that have been previously formalized to a (t), or stated in the form of an algorithm. An algorithm is defined as a precise and unambiguous statement of all the operations comprised in a process and of their sequence. Hence, the first thing to do is to learn to make formal statements of the complex thought processes that occur during document analysis, abstracting, indexing and retrieval. This is an important general problem which by far surpasses the scope of scientific information activities and informatics.

We have spoken quite at length about the relative merits and shortcomings of conventional and descriptor systems. Some believe that the former type of systems are difficult to mechanize while the latter are easy to mechanize. This opinion is not exactly true. Conventional systems, in particular those based on UDC, can be mechanized, too; but the point is that it is not necessary since they are efficient enough when embodied as a card catalogue. The case is different with descriptor systems - mechanized or automated retrieval is to some extent imperative with them, since manual searches for descriptor combinations are feasible only in relatively small files.

Material information carriers

A broad variety of technical facilities - ranging from the simplest information retrieval devices up to the latest - generation electronic computers - are currently applied to mechanize and automate information retrieval. Pursuing the objectives stated at the outset, we shall not go into detail but concentrate on some simpler devices. Still, obtaining a general notion of the whole gamut of equipment applied seems desirable. Usually information retrieval devices are classified by the information carriers they use. These are conventionally divided into discrete and continuous carriers. Here is a possible classification of the principal information carriers:

Discrete information media

Catalogue cards

Uniterm cards

Punched cards

---

1. Brillouin L. Scientific Uncertainty, and Information. New York-London, Academic Press, 1964, p. 39.

- hand-sorted
    - edge-punched
    - body-punched
    - superimposable
- mechanically sorted

Microfiches

Continuous information media

- Microfilm
- Punched tape
- Magnetic tape (disks, drums)

Catalogue cards and Uniterm cards have been described in the preceding lectures, so we shall rightaway come down to punched cards.

Punched cards

A punched card is a stiff paper rectangle with either holes punched along its edges or locations for punching such calibrated holes marked over the whole card surface. Punched cards were first applied in the 1880's by J.M. Jacquard who used them to control the loom of his invention. Early in the 1890's the American H. Hollerith introduced a pantograph punch and an electric accounting tabulator for body-punched cards which he offered for processing the U.S. Census data. This particular type of punched cards is therefore sometimes called "Hollerith cards". He founded a firm which later grew into a major computer company - "International Business Machines (IBM)."

The division of punched cards into hand- and machine-sorted is merely conventional since machines can be used to handle all kinds of cards, though edge- and body-punched cards as well as superimposable cards are as a regular rule processed manually. It should be emphasized that punched cards as well as all the other information carriers were not created expressly for information retrieval; they are used in record-keeping, statistics and for some other purposes.

Edge-punched cards are used in document retrieval systems with serial organization of AS. One or more rows of holes are punched along the card edges. Each hole is a binary code position permitting to code the presence or absence of a characteristic. The presence of a characteristic is indicated by clipping out the partition separating the hole from the card edge. The only thing to do in order to select the cards having a notch in a specified position (that is the cards on which the respective characteristic is recorded) is to insert a sorting needle in a card pack and shake it slightly. The cards with the notch will drop out.

Edge-punched cards of different sizes and with varying numbers of positions are produced. For their visual alignment in the file, the right-hand upper corner of every card is cut off. Most popular are cards sized 207 x 147 mm with two rows of holes (a total of 215 positions of which 200 are used). Various types of auxiliary equipment are used to manipulate such cards. These include tongs for edge-notching individual cards, keyboard devices - both hand - and electrically driven - for notching card packs, sorting-needles, and diverse sorting boxes and selectors.

The most important parameter of edge-punched card is the maximum number of different characteristics it can accommodate. Since the characteristics are recorded on the card by their codes the totality of holes along the edge is called a card's code field. The information capacity of a code field, or the number of different characteristics it can accommodate, is determined by the number of holes and the type of code applied. According to the method of representing the code words on an edge-punched card, a code may be direct or combined; according to its designation - selector or sequence; according to the type of encoded symbols - numerical or alphabetical; and according to positioning of the records in the code field - local or superimposed.

The direct code puts every characteristic into one-to-one correspondence with a hole in the card's code field. A direct code can be used to record as many different characteristics as there are positions in the code field. Advantages of a direct code are: no "retrieval noise" and minimal time and labour costs of encoding and card sorting. The drawbacks are the limited number of characteristics by which documents can be encoded and searched and the requirement that only mutually exclusive characteristics be encoded in every pair of holes in the case of a double-row card. The first shortcoming necessitates the application of combined codes - selector or sequence; the second shortcoming is common to all local codes, and to get rid of it a superimposed code is applied.

The selector code is a combined code that provides for retrieving the required cards from a file via a minimal number of sortings. A selector code puts every characteristic into one-to-one correspondence with a configuration of the code symbols recorded in a certain part of the code field (a subfield) dedicated to a given group of characteristics. Every word encoded by a selector code must comprise the same number of code symbols. Only one characteristic out of a given group may be encoded in a subfield whatever the combined code applied. This is a shortcoming of all local codes which makes them suitable for mutually exclusive characteristics only. A selector code is applicable both with single- and double-row cards.

The triangle (pyramid) code is a variety of selector code. It records every symbol with two notches in the positions situated at the upper ends of two diagonal columns of the coding network at the intersection of which the symbol to be coded is located. An advantage of the triangle codes as compared with linear ones is the smaller code field required. When a triangle code is applied to double-row cards the symbols are recorded with deep and shallow notches. The deep notches are made on the side on which the symbol in question lies at the intersection of the diagonal columns. In case the symbols are located one above the other the upper symbol corresponds to the right-hand deep notch.

The sequence code is a combined code providing for arranging the cards in a required order by a minimal number of sortings. With this code, every characteristic is put into one-to-one correspondence with a configuration of code symbols inside a code subfield. The configuration may consist of any number of symbols, provided they do not exceed in number the positions in the subfield. The sequence code has the common shortcoming of the local codes - it is fit for mutually exclusive characteristics only. But its advantage is that it takes less code field space than the direct and selector codes. The price of this, however, is greater retrieval noise due to more false-drops of cards with coincident notches.

All the above types of code may be either numerical or alphabetical. An alphabetical code may be regarded as a version of a numerical code since each letter of the alphabet is - explicitly or implicitly - uniquely designated by its ordinal number in the alphabet. When sequence code is used to represent letters their rapid alphabetic ordering is possible.

The superimposed code is a code meant to record more than one characteristic in one and the same code field. A superimposed code uniquely assigns to each characteristic a configuration of code symbols; the number of symbols per code word must be constant. An advantage of the superimposed code is a substantially increased utilization of the code field which is tantamount to increasing the latter's capacity. The fault with superimposed coding is the retrieval noise which can be caused by false configurations likely to arise when two true codes are superimposed. For example, if codes 17 and 23 are recorded in the same code field, false configurations 12, 13, 72 and 73 will appear. Some of these may coincide with true code configurations, which will result in retrieval noise. The noise will be the greater the more codes are recorded in the same field and the less symbols every code word consists of. The typical codes used with edge-punched cards are shown in Fig. 30. Codes used with body-punched and machine-sorted cards are essentially similar.

It should be stressed that edge-punched cards are recommendable only for mechanized information retrieval in files of up to 10,000 documents in the case of manual sorting and up to 30,000 documents with vibration or shock-type selectors. They are mainly applied to individual files in narrow fields of science and technology containing one or two thousand documents at most.

Body-punched cards are a modification of edge-punched cards and they are likewise intended for information retrieval systems with serial organization of AS. The code field of a body-punched card is shaped as a rectangular matrix of calibrated holes with equal spaces between the rows and the columns. A code position in a body-punched card is the partition between two adjacent holes either vertically or horizontally, depending on the card type. The coding of retrieval characteristics consists in cutting out a partition which results in a slot appearing in its place.

The searching is carried out in an unordered file. The cards are placed in the sorting box of a special selector. Sorting needles are inserted in the holes made in the box and in the cards, namely the holes bordering from above (or on the side) the required positions in the code field. Then the sorting box is tipped over to 180° or 90°. All the cards having slots in the corresponding positions will thrust out of the file to the extent of a slot. The box remaining in the tilted position, an auxiliary needle is inserted in a hole located at the uppermost (or sidemost) edge of the cards and the box. This needle connects the bulk of the file to the box except for the cards that have moved out (their edges were below the hole through which the connecting needle was inserted).

The selector box is then restored to its original position; but the auxiliary connecting needle prevents the cards that have moved out from dropping back into the file. The sorting needles can now be removed to take out of the file the cards whose code fields contain the sets of characteristics that constitute the required documents' search patterns. The chief advantage of body-punched cards lies in their being better fit for multi-aspect searching than edge-punched cards, that is for searching for several characteristics at a time. However,a special selector is needed with them. The working principle of body-punched cards is schematically shown in Fig. 31 (top).

Superimposable cards are used with inverted AS organization. This means that to every descriptor a special card is dedicated upon which identification numbers of all the documents having that descriptor in their search patterns are recorded. The recording is effected by punching a hole in the location whose coordinates indicate the document's call number. During retrieval the cards dedicated to all the descriptors of the search request are extracted from the file, laid one upon another and inspected against a light source to spot the coincident holes. The numbers indicated by the coincident holes are the identification numbers of the documents sought. Thus a superimposable card is in a way a modification of the Uniterm card permitting mechanized identification of the matching numbers in the cards selected for searching. The working principle of superimposable cards is schematically shown in Fig. 31 (bottom).

The main shortcoming of these cards is their limited information capacity which depends on the density of holes and the reduction of the space occupied by one hole. The card sizes vary between 187 x 82.5 mm and 445 x 445 mm, with capacities ranging from 400 to 400,000 positions. They are usually most efficient with files of up to 100,000 documents. To facilitate punching, a grid of coordinates is usually printed on cards, otherwise transparent stencils are applied. Punching or drilling devices are used to make the holes, and light screens, to spot the matching holes. Superimposable cards are either filed according to strictly alphabetized descriptors or sometimes they are provided with edge notches enabling them to be filed randomly. It should be noted that superimposable cards like Uniterm cards can be used to build grammarless descriptor languages.

Superimposable cards were invented, as mentioned, in 1915 by American ornithologist H. Taylor who applied them to define the different bird species by combinations of their characteristic features. It was however, not earlier than the 1940's that they were first applied to information retrieval. Superimposable cards are sometimes also called optical coincidence cards, feature cards and peek-a-boo cards.

Mechanically sorted cards are a variety of body-punched cards. They are of standard size - 187 x 82.5 mm. The face side of a machine card is covered with columns of figures printed over the whole field leaving free but a narrow horizontal band at the top. The figures constitute a matrix. The 45- , 80- and 90- column cards are the most common types in use. In contrast with hand-sorted cards, they have the left-hand upper corner cut off. Mass-scale-produced comparatively inexpensive computing and punching machines are used to handle machine cards. Systems based on this equipment can handle files of over 200,000 documents, and even several million documents, provided a preliminary ordering (subsorting). The equipment is efficient, simple and reliable. Both serial and inverted organizations of AS may be applied to such mechanized IR systems. The former type of systems are mostly built around sorting machines, while the latter type are based on card-selecting equipment.

In the case of a serial organization of AS, every card carries an encoded search pattern of a document and the letter's identification number. Sometimes the bibliographic details of the document and even its annotation are printed on every card, too. During retrieval the sorting machine scans the file or some part of it matching the codes of document search patterns recorded on the cards against the search request code entered into it. When the adopted matching criterion is fulfilled, the sorter selects the relevant card and conveys it to the receiver, for this card indicates the identification number of the document (and, in some systems, other data as well, which permits to regulate the searching procedure by the intermediate retrieval results). In the case of an inverted organization of the AS, every card carries the code of a descriptor and the identification number of a document having that descriptor in its search pattern. The rest of the code field may be used for a bibliographic entry and annotation which serve the same purpose as with serial storage organization. In the course of retrieval, subfiles of cards corresponding to the descriptors of the search request are selected. (Mark that whatever the procedure, standard card selecting equipment requires that a separate card be reserved for every document). In case an ordered file is used, this selection is done manually. A sorting machine is used to select the required descriptor cards from an unordered file.

Next, the selected card subfiles are ordered by their numbers and entered into a card-selecting machine, which selects out of the first pair of subfiles being processed only such cards as contain matching numbers. The resulting subfile is then matched against the third subfile etc., until the cards with matching identification numbers have been selected from all the subfiles. These numbers indicate the documents sought. It should be mentioned that despite the fact that a card sorter has an operating speed of 250 card per minute, the advantage of inverted over serial AS organization is practically reduced to zero by the magnification of the card file it necessitates.

Modified card sorters are now under development that will be able to handle cards carrying dozens of numbers. This will cut down the file size and hence the retrieval time.

Thus far we have made only some cursory remarks on the contents of the card space lying outside the code field. In some cards - edge-punched, body-punched, and mechanically sorted - this space may be sufficient to hold the identification number and bibliographic details of a document and even an informative abstract. Moreover, there exists a special type of card having a calibrated window (aperture) for inserting a microfilmed copy of a document. These cards are called aperture cards. They are also sometimes called Filmsort cards by the name of the company that first produced them on a commercial basis. With various types of cards the aperture size may be between 105 x 77 mm and to 30 x 41 mm. Most widespread are mechanically sorted 80-column aperture cards enclosing a frame of 35-mm film. Different types of aperture cards are shown in Fig. 32.

The chief advantage of the aperture cards is an 80 per cent reduction of document storage space as well as the facility of rapid retrieval and reproduction of documents. Usually they are applied in single-circuit IR systems in which AS and PS are combined in a single file of unordered aperture cards. Such systems are particularly efficient for storing technical drawings - a case where space and speed problems are especially acute. Suffice it to say that in designing a jet aircraft tens of thousands of drawings are used weighing sometimes as much as the aircraft itself. Essentially, aperture cards are a combination of a punched card and another information carrier, a microfiche.

Microfiches

Transparent microcards, or microfiches are pieces of negative or positive photographic film carrying the images of documents or their parts. They should be distinguished from opaque microcards which are made of an opaque material, usually photographic paper. Transparent microcards are sometimes called microfiches. They are of different sizes, but the currently most used size is 75 x 125 mm, that is the size of a standard library card. The upper horizontal band of a transparent card is reserved for the bibliographic reference readable with the naked eye. The document text is micro-reproduced (depending on the reduction ratio one transparent microcard may accommodate from 48 to 200 images) and can be read by means of special equipment. Specimens of transparent microcards are shown on Fig. 33.

The fault with single-circuit mechanically sorted aperture card IR systems is that the film is soon ruined due to scratching during sorting. For that reason systems with separate AS and PS are much more practical. It is efficient to combine AS implemented on usual mechanically sorted cards with PS consisting of hand-ordered transparent microcards. The merits of transparent cards - simple and easy filing, storage and production of enlarged copies - make them a most promising information carrier for PS in systems numbering millions of items.

Very sophisticated IR systems with speciall designed retrieval devices are developed on the basis of transparent microcards. An early system of that kind was "Filmorex" built by French physician J. Samain in 1950. It has been considerably modified and upgraded since. The information carrier used is a 35 x 60 mm-sized transparent microcard, shown in Fig. 33. It is divided into two zones, one of which is for the micro-image of the document and the other one for its search pattern. The maximum code field capacity is 20 descriptors, one vertical column per descriptor. Filmorex system comprises the following hardware: a recording camera, a guillotine for cutting film into frames, a selector, a reader-printer, and card files - subject, auxiliary and author.

Special code cards are employed to record document search patterns on microfiches. For every descriptor there is one such card carrying its five-digit number and the binary code for that number. The operator of the recording camera pulls out of the file the code cards of the descriptors of the document being processed and lays them one upon another in such a fashion that each subsequent card covers the underlying card but leaves open the edge with the binary code of the descriptor. This superposition forms the code of the document search pattern and it is photographed together with the document. The last (topmost) card remains entirely open, and thus, in addition to the binary code of that descriptor its numeric, alphanumeric or alphabetic code is also photographed. On the transparent microcard, this code reproduced beside the code field, is easily readable with the unaided eye. The basic advantages of this code-forming technique are its simplicity and the fact that it is almost completely error-free.

During input, each document is photographed as many times as there are descriptors in its search pattern; the last, uppermost code card is every time taken off and placed under the rest of the cards so that it becomes the bottom one. As the result the code of the next descriptor is laid open and recorded on the next transparent microcard. Thus every following card is denoted by the code of the subsequent descriptor. The cards are then manually sorted according to these codes between the corresponding sections of the operational files. It takes an average of 2 minutes to enter a document with a search pattern of 5 descriptors.

Retrieval is performed with the help of a special selector. The search request is recorded in a complementary code, that is in such a way that the positions which are transparent in the basic descriptor code are black in the complementary code and vice versa. When superimposing a complementary code upon the corresponding basic code record, the line of the code field becomes completely masked out. Cards can be searched by three descriptors at a time at most. For that, a "negative" mask of the codes of the descriptors is prepared; this mask is a piece of black film into which the complementary codes of the descriptors have been punched on a keyboard punch.

The mask is entered into the selector and during the sorting of the cards is superimposed, as it were, on their code fields. If a mask completely blacks out the code field of a microcard this means that this field contains the required descriptor codes. The coincidence of a code with its "negative counterpart" is established by three light cells that close the electrical circuit of the sorter magnet which transports the microcard to the selector holder. The output device of the system is a reader-printer which produces enlarged document copies ready for use. The selector can handle up to 600 cards in one minute, and the reader-printer up to 40 cards per minute. The operation and the principal units of the Filmorex systems are shown in Fig. 34. Its cost is on the order of $25,000. Several dozen such systems are currently operational in a number of countries.

Other, more complex systems based on transparent microcards have been developed as well, for example Minicard, Magnacard, Walnut, costing from hundreds of thousands to millions of dollars. However, these expensive systems exist but as single specimens and have not become widespread. As a matter of fact, the growing popularity of transparent microcards is explained just by their being efficiently applicable without any complex and costly equipment.

Microfilm

The same is nearly as true of microfilm, which,though, belongs to continuous information carriers. The division of material information carriers into discrete and continuous is as conventional as any other classification. Indeed, if we cut microfilm into small strips of several frames (called strip microfilm) they can be viewed as discrete carriers. For that matter, even roll films are continuous only within a roll. The continuous nature of a carrier creates certain difficulties with ordering the file of documents it stores. On the other hand, equipment used with continuous carriers ensures very high speeds of operation. The evident advantage of microfilm is that it requires no special film-handling equipment, since all the existing inventory of devices of the film industry is directly applicable.

Still, information retrieval of microfilmed documents does require special microfilm selectors. The concept of such selectors was first suggested by German E. Goldberg back in 1927. The first operational prototype microfilm selector "Memex" was built in the late 1930's by V. Bush of the Massachusetts Institute of Technology whose words have been quoted at the beginning of this lecture. The final version of the device, called "Rapid Selector", was constructed by R. Show in 1949; since then it has been repeatedly modified.

The Rapid Selector handles 35-mm unperforated positive film. Recording of documents is carried out at reduction ratios from 2.5:1 to 20:1. The frame size is 29 x 44 mm; every frame on the film is preceded by a code field for a document search pattern. The capacity of the code field is 315 positions. The search pattern is entered into the code field from an 80-column card punched on a keyboard punch. The punched card is fed into the microfilming camera, and the punched holes are automatically converted into a binary light code which is projected on to the microfilm.

Search requests are likewise entered into the selector from 80-column punched cards. A switchboard panel is used to specify the logical relations between the search pattern descriptors. As the film moves a special photoelectronic unit matches the search request against the search patterns of the documents whose microrecords are on the film scanned. The documents produced by the selector are copied onto the reproduction microfilm in a dynamic mode, that is without stopping the main film. At the film drawing speed of 2.7 m/sec and a main reel capacity of 1,800 m, the mean retrieval time in a reel, including development of the reproduction film, does not exceed 12 minutes.

A "Poisk-OK" microfilm selector has been developed in the Soviet Union in 1962 by engineer Yu. Ya. Klyachkin (VINITI) for searching documents' micro-records by their identification numbers. The information carrier used is again 35-mm unperforated positive film with documents filmed at a reduction of 15:1. Identification numbers are entered from a numeric keyboard located on the console of the filming device. The number is projected against a coding scale and recorded together with the document. On the film, the document (double list) and its number are accommodated in a frame sized 31 x 43.5 mm. The density of filmed records is 19 frames/m. The code field with a document's call number precedes every frame containing a copy of that document. 250-270 m reels are used accommodating up to 5,000 frames each.

Requested numbers are keyed into the selector and the identified images are projected onto the display screen enlarged to the original size. The mean retrieval time in one reel (holding up to 10,000 pages) at a film-moving speed of 0.6 m/min. does not exceed 4 minutes. There is a facility for stopping the film to make natural-size copies of the images from the display on usual paper. (Hence the name of this model - "Poisk-OK" which means "Searching with stop for copying"). A new model of this selector has now been developed, named "Poisk-DV" (which stands for "Searching with dynamic copying"). While retaining the main features of the previous model, the new version permits to copy the requested images onto 35-mm roll film without stopping the film drive. Atomatic control of the selector is built in.

There exist still more complex selectors with better performance characteristics. These include, among others, US-manufactured equipment used in IR systems MIRACODE, RADIR, CRIS, FLIP. It should be noted, however, that microfilm selectors are mostly used as PS in IR systems with a relatively limited file. This explains the marked trend towards simplification of these selectors. Some recent popular models can even be regarded as ordinary reader-printers supplemented with elementary facilities for searching filmed images and a mechanical film drive.

Roll microfilm plus simple and inexpensive readers and copiers can well be used by individuals for building personal scientific and technical libraries.

Electronic computer and automated information retrieval

Of all the continuous information carriers listed above we have not yet spoken only of punched paper tape and magnetic tape, disks and drums. Punched tapes are much used in devices employed to compile various information publications-permuted-title indexes and union catalogues. The particular types of these devices will be discussed in the following lecture. Magnetic tape is applied in some special-purpose information retrieval devices, for example those used in the Videofile system (USA, 1958-1964). These devices will by all likeness have good prospects after television techniques take root in scientific information activities. Meanwhile, however, they are rather rarely used due to their complex design and high costs, and for that reason we shall not dwell on them.

Continuous carriers of the above types are mainly used for input to, and external memories of, electronic computers, which are more and more regularly applied to information retrieval tasks. The design and operation of the different types of computer would make the subject of a different and quite an ample course. Still, discussing mechanization and automation of information retrieval we cannot dispense with assessing the part the electronic computer will have to play in this sphere. Successful applications of the computer in the various fields of human activity have inspired a great number of people with what almost amounts to a mystical faith in its infallibility and omnipotence. However, the real capabilities of computers are far from boundless.

The essential advantages that a computer has over human brain are uniformity of operations, reliability, minimal errors, high speed, and eventually lesser cost of certain operations. All these davantages can be realized only provided there is an algorithm and a program for the performance of the task concerned. In contrast to man, the computer will always give identical indexing to one and the same document, and it can be by far faster and more exact in comparing an information request against the abstracts of documents or even their complete texts.

However, unless we understand the inner mechanism of such thought processes as document content analysis, abstracting, annotating, and indexing, we shall not be able to algorithmize the processes. This, in turn, requires deep-going studies into thought psychology, linguistics, mathematical logic, semiotics and other disciplines. All attempts at using computers for information retrieval todate have been based on mere statistical properties of document texts and did not take into account their semantic properties. But we are not even certain that the identity of meaning of two texts can be established by purely statistical methods without any semantic analysis of their contents that involves a reference to extra-linguistic information, that is information not immediately contained in the texts being processed.

The greatest promise therefore appears to lie with finding the optimal modes of man-machine interaction so that their respective advantages could be utilized and the one supported the other in the best possible way. As Norbert Wiener, the father of cybernetics, phrased it in his last work: "Render unto man the things which are man's and unto the computer the things which are the computer's. This would seem the intelligent policy to adopt when we employ men and computers together in common undertakings. It is a policy as far removed from that of the gadget worshipper as it is from the man who sees only blasphemy and the degradation of man in the use of any mechanical adjuvants whatever to thoughts".(1)

To recapitulate. We have outlined in general terms the technical facilities of information retrieval. In keeping with our chief objective - just to usher a student of informatics into the science - we concentrated on the aspects essential for mechanization and automation of information retrieval. As to particular information-retrieval media, the emphasis was on those using punched

---

1. Wiener N. God and Golem, Inc. A Comment on Certain Points where Cybernetics Impinges on Religion. Cambridge, The M.I.T. Press, 1964, p. 72-73.

cards, which can be considered the most generally available and fairly efficient under certain conditions. At shorter length we have spoken of the devices based on microrecords of document texts, despite the fact that these devices have a major part to play in mechanizing scientific information activities. To some extent this gap will be eked out by the following lecture where the copying and reproduction of science documents will be dealt with as the starting stage and essential prerequisite to the utilization of scientific information.

Checkup questions

1. Under what conditions is it advisable to mechanize and automate scientific information activities?
2. What are the existing information carriers and how are they applied?
3. What are the applications of the different types of punched cards?
4. What are the advantages and shortcomings of the principal punched card codes?
5. What are the principles of handling edge-punched, body-punched, and superimposable cards?
6. What are microfilm selectors used for?
7. How are computers to contribute to the solution of information retrieval problems?

Literature

1. Becker J., Hayes R.M. Information Storage and Retrieval: Tools, Elements, Theories, New York, Wiley, 1963.
2. Punched Cards. Their Applications to Science and Industry. Ed. by R.S. Casey, J.W. Perry, M.M. Berry, A. Kent. 2nd ed. New York, Reinhold, 1959.
3. Scheele M. Punch-Card Methods in Research and Documentation. (With special reference to biology). New York, Interscience Publishers, 1957.
4. Introduction to Data Processing.- "Library Resources & Technical Services", 1965, vol. 9, N 1, p. 5-103.
5. Taube M. Computers and Common Sense. The Myth of Thinking Machines. New York-London, Columbia University Press, 1961.

9. REPRODUCTION OF SCIENTIFIC DOCUMENTS

During the entire course we have repeatedly stressed the fact that a scientific document is a major attribute of scientific information work. What is more, the scientific documents are a mode of existence of science, for they are the end product of any research, they serve as one of the primary means of dissemination of scientific information, and their utilization poses most important organizational problems of scientific research.

For this reason document reproduction, i.e. copying and duplicating, cannot be viewed as purely technical matter. The level of reproduction techniques and facilities will have a direct bearing on the effectiveness of scientific information utilization and, eventually, on the unhampered advancement of science. For centuries since Gutenberg's invention, the typographic process has remained the sole method of edition printing of textual documents. It was only in the recent decades that photography, invented in the 1830's, found extensive application as a document copying method.

Document reproduction processes

The need for speedy and low-cost copying techniques and facilities, which arose several decades ago in the sphere of clerical and office work, has promoted the development of a variety of copying and duplicating processes. In addition to the former silver halide and diazo processes (i.e. photography and dyeline copying), the thermographic and electrographic processes which first appeared in the early '50's have quickly made their way into the document reproduction field. It is now beyond doubt that the fast spread of new reproduction techniques has no smaller role to play in enhancing the efficiency of scientific research than the invention of book printing five hundred years ago. The theoretical and practical aspects of these processes are dealt with by a new technical discipline which has been termed reprography. Familiarity with the essentials of reprography is important both for the information officers and for subject specialists in the different areas of research and practical activities. The development and use of the various information retrieval devices which were discussed in the preceding lecture, is only made possible through the application of the up-to-date copying and duplication facilities.

Document reproduction processes can be broadly classified as follows.

Document copying

- Silver halide processes
    - Photographic
    - Reflex
- Diazo processes (dyeline copying)
- Thermographic processes
- Electrographic processes

Document duplicating

- Typographic printing
    - Letterpress
    - Cravure
    - Flat-bed (lithographic)
- Spirit duplicating
- Stencil duplicating
- Offset duplicating
    - Addressing machines

Division of all document reproduction processes into copying and duplicating is for the sake of convenience only; the criteria of division are the number of copies produced and the use of a specially prepared master. Copying is generally held to include those processes which yield a maximum of 12 or 15 copies from any original which has not been specially prepared for this purpose. But, then, many copying processes will be able to produce more than 15 copies, and besides, they place certain demands on the originals being reproduced.

Photography

Copying of documents by means of photography was ushered in by the invention of photography. Back in 1839, when J.M. Daguerre had for the first time demonstrated his photographic process to the members of the French Académie des Sciences, a German medical student at the University of Liège, by name of Albrecht Breyer, began experimenting with silver halide-coated paper to produce copies from printed book pages. At present, standard photographic processes are used to obtain high-quality copies enduring prolonged archival storage. These processes are, however, very costly and time-consuming and require specially equipped darkrooms. Standard photographic processes are mainly employed in scientific information work for producing copies of half-tone illustrations with further duplication in multiple copies (by the photostatic method), as well as for producing reduced copies (microcopying) or enlarged copies (reproduction copying).

Production of greatly reduced copies of documents is called microphotographing or microfilming. Microfilming (with reduction ratios of from 8 to 200 times) permits to reduce for storage the size of the original documents by as much as 95%, to standardize the document size, to promptly execute copying and small-run duplicating jobs. Microcopy negatives serve as starting materials for the "same-size" reproduction of originals. Microcopy positives are used directly for reading by means of microfilm readers.

The microfilm readers may have reflexive or translucent screen. Commercially available are stationary, desk-top-type and portable readers which offer a magnification range from 8 to 200 times. In most of them microphotographs are read against a light source. To obtain enlarged copies of microfilm, both standard enlargers and special electrographic apparatus can be used. Steadily increasing use is made of the reader-printer devices which make it possible, after reading a micro-frame, in a roll film or microcard, to obtain immediately an enlarged copy of this frame on special paper.

Reflex processes

To overcome the difficulties attending the standard photographic processes, to facilitate silver halide copying processes and to eliminate the need for darkrooms, the reflex processes have been developed. They are based on the contact copying method, where the exposure of the sensitized paper is effected by the light reflected from the surface of the original being copied. The most widespread are the stabilization, diffusion-transfer-reversal and gelatin-dye-transfer processes.

In the stabilization process, silver halide-coated paper is laid with the coating on the original being copied and exposed to light from the side of the copy paper. The light passing through the sensitized material is absorbed by the dark areas of the original, and reflected from the light areas back onto the coating. This results in the reflex exposure of the copy paper. The paper is dipped into an activating solution which converts the exposed silver halide to metallic silver in the areas corresponding to the non-image portions of the original. After development the copy paper is put into a stabilizing solution the constituents of which (thiourea) react with the unexposed silver halide to form complex compounds which are insensitive to light. The next step is thorough squeezing and drying of the paper. The first reflex copy obtained is negative. Using it as an original, further re-runs produce any wanted number of positive copies. It should be noted that the need for drying of the original renders the stabilization process too slow (though less so than standard photography which requires time-consuming washing after fixing).

The diffusion-transfer-reversal (DTR) process is much faster, but requires two types of special paper, negative and positive. Negative paper is coated with gelatin emulsion which contains silver halide and other light sensitive substances. Positive paper has a coating of colloidal silver and is insensitive to light. Negative paper is exposed by the reflex method, as in the previous process. Following exposure the negative paper is removed from the original and immediately, without preliminary development, is placed face to face with the positive paper. The two are submerged into a solution containing a developer (e.g. methol and hydrochinone) and a solvent for silver halide (e.g. sodium thiosulfate). In the areas of the latent image of the negative, soluble silver salts are formed which are transferred by diffusion to the positive paper, where under the effect of colloidal silver, they form metallic silver. The positive image is thus produced. DTR is the quickest of the reflex processes. However, it has also some drawbacks, notably one negative can produce only one copy, on which brownish spots may appear and which may become yellow with time.

The gelatin-dye-transfer (known under the trade name Verifax) depends on the use of a printing matrix, which is a sheet of light sensitive paper coated with a gelatin emulsion containing silver halide, a non-diffusing developing agent and a dye forming agent. Exposure of the matrix is by the standard reflex method. It is then placed in a solution of alkali known as the activator solution which has the effect of hardening the gelatin on the lighted areas, while the dye forming agent converts to unsoluble blue dye. This dye can be transferred onto a sheet of any paper together with the unhardened gelatin in which it is contained. The gelatin which has been transferred to paper contains not only dye but also silver halide which under the action of thiourea (with which the paper has been impregnated previously) gradually converts to metallic silver. After the dye has faded out this metallic silver produces the copy image. One matrix will yield up to seven copies. Since the sensitivity to light of the gelatin coating on the matrix is fairly low, copying can be carried out in a room with subdued fluorescent lighting. The matrix is sensitive, though, to the yellow colour. As in other reflex processes, yellow, orange and green originals cannot be copied. Besides, the gelatin-dye-transfer process does not produce quality copies from half-tone originals and requires specially trained operators. The schematic of the reflex and other copying processes is shown in Fig. 35.

Diazo

The diazo process, also called dyeline and diazotypy, is based on the fact that many diazo compounds entering into reaction with aromatic amines, phenols and some other compounds, form azo dyes. But, diazo compounds are decomposed when exposed to ultraviolet light and become incapable of forming azo dyes. For copying, an original is placed on diazo paper or film and exposed to an ultraviolet light source from the side of the original. This results in the decomposition of the diazo compound in all areas of the coating which are not screened by the elements of the image being copied. The ability to form azo dyes is retained by those areas of the diazo coating which correspond to the elements of the image being copied. If, therefore, such paper or film is subjected to appropriate chemical treatment, a positive copy of the original will appear on it.

According to the development method, diazo processes are divided into dry and wet. In the dry method the diazo coating contains azo components which react with non-decomposed diazo compound to produce the corresponding dye. In addition, such a diazo coating contains an acid preventing premature coupling of diazo compound with the azo component. The exposed diazo coating is brought in contact with ammonia vapour which neutralizes the acid, resulting in the necessary reaction which yields the dye. In the wet method the light-sensitive coating contains only a diazo compound. Because of this fact, the diazo paper or film is treated with an alkaline solution containing a certain azo component.

The dyeline process is used primarily for the duplication of engineering drawings. It has come into wide use since the middle of the last century, after John Herschel, a British chemist and astronomer, invented a light-sensitive paper which originally was cyanotype. Diazocopies retain sharp image for some 25 years, and their cost is four or five times lower than that of the copies produced by any other method. However, the copied original must be on one side of a transparent or translucent material.

Thermography

Thermo-Fax, is the most popular of the thermographic processes. It is direct, i.e. it involves no image transfer. It was introduced in 1950 by the Minnesota Mining and Manufacturing Co. (3 M) as a result of the investigations by a physical chemist, Dr. Carl S. Miller. This process makes use of heat-sensitive paper which contains compounds that form a coloured substance when heated. The reaction usually involves a ferric compound, such as ferric stearite, and a phenolic compound, such as tannic acid. These materials are held in a binder, such as polyvinyl butiral. To prevent premature reaction between these compounds a waxy ferric compound is used. The reaction which forms a coloured product will take place only after the ferric compound has melted.

For copying, heat-sensitive paper is placed face up on the surface of the original to be copied and is exposed to infrared radiation from the side of the copy paper. Infrared passes through the heat-sensitive paper and, being absorbed by the printed areas of the original, generates heat that is transmitted to the corresponding copy paper areas. This absorption of infrared occurs only when the image being copied is printed in inks which contain carbon black or metallic compound. Organic inks (of vegetable origin) do not absorb infrared. As a result of heating the ferric compound converts to a melted form and, reacting with the phenolic compound, produces a positive copy of the original on the heat-sensitive paper.

Thermography is applied to the copying of printed or typed originals, as well as of pencil inscriptions and drawings (pen and ball-point inscriptions cannot be copied), where it is necessary to produce copies for temporary use. This is a singularly easy and cheap process, but the copies it produces will become blackened and brittle especially if exposed to heat. There are also some indirect thermographic processes, e.g. the Dual Spectrum process, or the distillation process, which overcome the shortcomings of thermography but at the same time slow down copying and render it much costlier and more complicated.

Electrography

Electrography is a name which is commonly applied to the various photoelectrical and electromagnetic processes of forming images and printing without impression. As early as 1921, a Russian inventor, E.E. Gorin, proposed to use the properties of photoconductors for image reproduction and filed an application for the invention of what he called an "Electrographic apparatus". Apparently, he was the first to use the term "electrography". The American physicist C.F. Carlson combined the principle of image forming which had been proposed by Gorin with the principle of development of latent electrostatic image by the sprinkling of a dye, proposed by P. Selenyi, a hungarian, and in 1938 obtained the first electrographic copy. This process was subsequently perfected at the Battelle Memorial Institute (USA) and re-named "xerography" (from the Greek words meaning "dry writing"). Commercial exploitation of electrography began in 1950. Electrography divides into processes involving image transfer (xerography) and those without image transfer (electrofax). Both are presented schematically in Fig.35.

Xerography works in the following way. A selenium coating on an aluminium plate is used as the photoconductor. To make the plate sensitive to light it is given a positive electrostatic charge by a corona discharge in the dark. During exposure of a charged plate in a reproducing or projecting apparatus the lighted areas of the selenium coating are depolarized. This results in the formation of latent electrostatic image on the surface of the photoconductor coating, for the development of which an oppositely charged toner is cascaded across the plate. The toner particles are attracted by the charged areas of the coating and held by them permanently. Fine synthetic or natural resin powder with the addition of a pigment serve as the toner. For electricizing the toner the so-called carriers are used which are small glass or polystirene beads.

A sheet of common paper is then laid over the powder image thus formed and given an electric charge by means of a corona discharge of the same sign as the charge of the latent image and opposite to the charge of the powder image. As a result, the toner particles are attracted to the paper surface and settle on it. The powder image is fixed on the paper surface either by heating the toner to its melting point or by exposing it to the vapours of a strong solvent, such as acetone or sulphuric ether.

An electrographic copy does not take more than 3 or 4 minutes to produce, even when all the operations are manual. At present, fully automatic rotary-type apparatus are used, in which the selenium coating is applied to a rotatable drum and not to a plate. Such apparatus has high operating speeds that enable a copy to be produced in a few seconds. Xerography uses a dry method which requires no darkroom, and the materials are readily available and fairly inexpensive. An image can be copied not only on paper, but also on fabric, films and even on metal masters which are used in offset duplicating. The chief drawback of xerography is insufficient quality in reproducing half-tone images and large solid areas. Besides, the equipment, particularly the rotary automatic machines, is very expensive. A selenium-coated plate can produce not more than 2,000 copies, and a drum must be replaced after about 50,000 copies.

The electrofax method, or direct electrostatic process, was introduced by the Radio Corporation of America (USA) in 1954. It makes use of photoconductor paper coated with zinc oxide on which electrographic copies of the original are directly formed. Electrographic paper is produced by applying to a paper sheet a coating of a dielectrical binder, such as silicone resin or polystirene, as well as organic dyes that sensitize zinc oxide to light. These properties of zinc oxide have been investigated by Soviet researchers Academician A.N. Terenin and E.K. Putseyko. Electrostatic paper is charged and exposed in the same manner as in the xerographic process. The only distinction is that here the paper is given a positive charge and the pigment, a negative one.

The visible image is produced on electrographic paper by either a dry or wet method. Dry development makes use of the "magnetic brush" technique. The developing powder contains the carrier in the shape of fine iron particles which, together with the charged pigmented particles that cling to them, are attracted by an electric magnet forming a sort of "brush". As this brush moves across the electrographic paper with latent image the pigmented powder adheres to the oppositely charged areas of the photoconductor coating and develops this image. The iron particles remain on the magnet. This method makes a fine job of line drawings and half-tone images, as well as large solid areas. The powder image is fixed on the electrographic paper by heating the pigment to its melting point. Liquid development is either by dipping the paper in the developer or by spraying the developer on the paper. The developer is usually a high-specific-resistance liquid (e.g. aircraft gasoline) in which the appropriate pigment is dispersed. It requires no further fixing of the image on the copy.

It should be noted that the Electrofax method permits to design high-speed automatic copying equipment, dispenses with limitations on the originals which can be copied, but it requires special paper from which the images can be easily scratched off by metal objects. In addition to the electrographic processes discussed, a few others are used, including the electrolytic and thermoplastic processes. We have considered only the basic types. Their relative merits and demerits are tabulated in Fig. 36.

It was our intention to leave out of discussion the specific copying devices. Such equipment includes thousands of models varying in complexity from the simplest, designed for manual operation, all the way to the most sophisticated apparatus designed for semi-automatic or fully automatic operation. Directly governed by their design and application are their speed, labour requirements and cost. Apart from that, in this new and fast growing field the new equipments are constantly entering the market. It can be stated with certainty that by the time you complete your course, the existing equipment will have been replaced by the new equipment having better performance indices. This explains why our introductory course is confined to giving a general overview of the novel copying processes. Once you understand them you have a key to mastering any other apparatus of the sort. Concluding this survey we should like only to point it out that electrography is becoming the most widespread of all the processes discussed, and seems to be the most promising of all the existing document reproduction methods.

Typographic printing

It has been said already that duplication is commonly understood to include all processes that produce in excess of 15 copies from a special master. The principal traditional method of document duplication is by typographic printing, which has been termed professional printing, or graphic arts, as distinct from the more modern duplication techniques.

In the fourth lecture, when speaking about the channels of dissemination of scientific information, we have touched upon the work of a publishing office and a printing shop. We discussed only the process of preparation of a form for relief printing. In actual fact, three kinds of printing can be distinguished: letter-press (relief), gravure, and flat-bed. In letterpress printing, the printing elements are raised above the rest of the form surface. The form in this case consists of a set, which reproduces the text, and plates which reproduce the illustrations. This printing method is used in the production of the great majority of books, periodicals, newspapers and other predominantly textual materials. In gravure printing, the printing elements are deepened with respect to the white space elements. The varying depths of the printing elements permit to reproduce gradual transition from the lighted to the darker areas on the original. This strongly recommends gravure printing for reproducing art illustrations. In flat-bed printing, all printing and non-printing elements are on a plane. This kind of printing (its most common variety is called lithography) is generally used for the multi-colour printing of maps, posters, post-cards etc. The different kinds of printing are schematically shown in Fig. 37.

The form is used in the printing of a run, i.e. in the production of the requisite number of copies. There are two techniques of doing it. With the first technique, the transfer of the ink from the form occurs at the very moment when the paper is pressed against the form. With the second technique, which is called offset printing, the ink from the form is transferred first to an intermediary roller, from which it is transferred on to the paper.

Modern printing shop equipment includes platen, cylinder and rotary machines. In platen machines the form and the surface which presses the paper (platen) are plane surfaces. In cylinder presses, only the form to which the paper is pressed by the impression cylinder has a plane surface. In rotary machines both the form, and the surface which presses against the paper are cylindrical. Substitution of cylinders for plane surfaces provides for a substantial speeding up of the printing machines. The working principles of the various types of printing machinery are demonstrated in Fig. 38.

Office reproduction techniques

The demand for facilities which would permit big-run duplication of documents, but faster, cheaper and easier than the conventional graphic arts methods has caused the emergence of a variety of processes broadly termed "office reproduction techniques". The processes include the means of non-composition master production, flat-bed offset, stencil and spirit duplicating, as well as advanced methods of editorial document processing. The emergence of office reproduction is a breakthrough in the information communication techniques which is comparable in impact with the invention of book printing. The merits of office reproduction - low cost, high speed, small size and general availability - permit almost the same quantitative and qualitative results as in graphic arts to be achieved with less labour, materials and time spending.

The advantages of office reproduction techniques over graphic arts are not confined to the reduction in the production cycle and, hence, publication time on account of the proof-reading step. Authors' and editorial corrections of the form are the essential steps in any typographic process, as they considerably improve the content of the published documents. However, office reproduction facilitates and speeds up the correction processes by employing the new means of master preparation. These means constitute the essential features of the new facilities and techniques of document duplicating.

Composing typewriters and phototypesetters

The majority of modern machines for the preparation of masters for office reproduction are traceable to the ordinary typewriters. Commercial manufacture of typewriters was started a century ago, in 1868, by the American military concern Remington. The typewriter design proposed by engineers K. Sholles and K. Gliden allowed printing in single-size letters. In 1878 the typewriter design was improved to accommodate both upper-case and lower-case letters. Further perfections, including the introduction of the tabulator, and the interline space regulator were introduced by another American concern, Underwood Typewriter Co., founded in 1895.

Owing to the introduction of carbon paper, in the late 1960's and early 1870's, the typewriters began to be used for the duplication of textual materials. In 1873, T.A. Edison invented a method of stencilling using a wax stencil, which marked the beginning of the use of typewriters as a means of master preparation. From the time the first typewriters entered the market and up till the present day their designs have been undergoing continuous changes. One of the most promising trends in this respect is the development of electric and automatic portable composing machines for master preparation. The first such machine was apparently the Skoropechatnik (literally meaning "Fast-Typer"), designed in 1873 by a Russian inventor, M.I. Alisov, and exhibited three years later in Philadelphia, where it was awarded a medal.

The present-day composing typewriters are distinguished by the great variety of type sizes, sets of type and styles of type-face, which closely resemble print type. These machines permit differential spacing according to the width of individual characters whereas in ordinary typewriters the spacing is the same for all characters and punctuation marks. Many composing machines provide line justification, which means that the right-hand edge of the printing master is an even line. The Varityper, which is one of the widely used machines of this kind, permits the composition of complex scientific and technical texts using mathematical, chemical and astronomical symbols. The Varityper belongs to the range of segment machines which have distinctive advantages over the usual machines. The characters are fixed on a changeable segment which is positioned on the type drum. Pressing a key turns the type drum to rotate until it faces the paper with the corresponding character, after which the impression takes place. The carbon-paper ribbon, which can be used only once, produces sharp and clear image which is suitable for the immediate reproduction or the preparation of offset masters. The Varityper can be operated at about 40,000 characters of mixed setting per working day, but with line justification, nearly twice slower, because each text has to be typed twice: during the first setting all the word spaces in a line are added up, and during the second setting the characters are spaced out and the line is thereby justified.

The Flexowriter composing typewriter produces simultaneously both a typescript copy which can be used for the preparation of a printing master, and a punched paper tape which can be used for the second - automatic - re-typing of the composed text. This tape can also serve as input to a computer. Because of this, Flexowriters are sometimes used as input devices with punched card equipment and electronic computers. A Flexowriter tape can be used for the preparation of permuted-title indexes, union catalogues and other information publications. The Justowriter composing machine effects semi-automatic typesetting with justification without re-typing. The Justowriter consists of two units called the Recorder and the Reproducer. The Recorder is a composing machine which produces both a typed copy in galley form and a punched tape complete with justification codes. The tape is inserted into the Reproducer which interpretes it and automatically produces a justified copy on ordinary paper or an offset master.

Along with composing typewriters, printing masters have been increasingly prepared by means of phototypesetting machines. These machines use a variety of type-face styles and signs of different alphabets and have great potentialities for changing type size. During composition, characters are projected on to a photographic paper of film, producing a negative or positive image of the text which, after proofing, correction and make-up, is copied on to a printing master. The widely known phototypesetter makes include Photon, Linofilm, Digiset and some others. The trend in the development of facilities for the preparation of printing masters and of office reproduction techniques generally is towards high-speed electronic printing devices permitting a direct output of textual and graphical information from computers.

The facilities described above permit to obtain only a starting original for the subsequent preparation of a master. Methods of master preparation vary with the type of printing used, viz. spirit, stencil or offset.

Spirit duplicating

In the preparation of masters for spirit duplicating, paper or other moisture-resistant material is brought into contact with a special carbon paper or ribbon covered with an alcohol-soluble dye. As the image is applied to the original, a laterally reversed image appears on the underside under the effect of the dye.

The spirit duplicating apparatus is called a hectograph. Its working principle is shown in Fig. 39 (top). The main parts of a hectograph are the impression drum (cylinder) on which the printing master is mounted, the pressure and leader rollers, the paper feeding and moistening devices. The paper feeding device feeds paper sheets,one at a time, into the apparatus. There the paper surface is uniformly damped by ethyl alcohol which dissolves the dye on the reverse side of the original serving as master, and transfers a small amount of the dye on to the imprint. The best imprints are obtained on writing paper which has optimal absorbency. Using other paper types will result in blurring and spreading of the image. By adjusting the damping rate and the pressure exerted upon the paper at the moment of contact with the master, it is possible to vary the brightness of the image and the number of imprints which can be obtained from one printing master. This number will vary within the range of 100 to 300 imprints of acceptable quality.

Now hectographs with replaceable stamping mechanism are available which permit to obtain imprints from masters to which modifications, additions, deletions and substitutions have been made. They are used for the production of multiple copies of forms with different filling in different columns, e.g. order, accounting, invoice forms, etc. These hectographs alternate printing from the fixed image master and from the stamping mechanism which contains data varying for each imprint.

Spirit duplicating is the only office reproduction process which can produce copies in up to six colours in one operation. This is due to the use of carbon paper with different colour dyes. The most bright and permanent image is provided by the crystal violet dyestuff. The imprints produced by spirit duplicating will fade with continued exposure to light and are also somewhat soluble in water. But these disadvantages are offset by the cheapness of the process and the fact that it does not require complex equipment and specially trained staff.

Stencil duplicating

The method of stencil duplicating consists in the forcing of a dye through a printing master, or stencil, on to paper surface. Preparation of the stencil involves the formation of printing elements in the shape of minute holes in the master through which the dye penetrates, and the formation of the white space elements which are impermeable by the dye.

Stencil duplicating is done by means of an office-type rotary stencil duplicator. It consists of the drum (cylinder) made of metal net and covered with fabric on the outside. On the outer surface of the drum the master is mounted. Inside the drum is the inking mechanism. As the drum rotates the ink roller soaked by the ink from a bath transfers the ink to the surface of the ink distributing roller from which it squeezes through the drum perforations on to the inking fabric and further on to copy paper. The imprint appears while the paper passes between the rotating drum and the impression roller due to the forcing of ink through the holes in the master. The working principle of a rotary press is shown in Fig. 39 (centre).

A stencil master, commonly called a "wax" stencil is a sheet of long fibre tissue impregnated with paraffin, wax or other ink-impervious substance. The stencil is inserted into a typewriter, and typing is performed without ribbon in order to displace or remove the ink-impervious coating. Because of the simplicity of the working cycle, and the low capital and operating costs of the equipment, stencil duplicating has found wide application in the production of small runs. However, this kind of printing has some shortcomings; notably it cannot cope with illustrative materials, the wax stencils will quickly deteriorate when exposed to heat, and the quality of text in the imprints is often of rather inferior quality.

Improvement in the quality of stencil duplicating is achieved by the use of the electrical spark method of preparation of the master. An electronic copying apparatus permits stencils to be produced from any printed, typed or inscribed original. The apparatus has two drums rotating at the same speed; on one is placed the original, on the other, the stencil. The light cell of the scanning device recognizes the textual images which are converted to high-frequency current and transmitted on to a special conductor rotofilm. The film is eroded by electrical sparks according to the tones of the areas being scanned. The stencil thus produced allows runs of up to several thousand copies to be made. Electrically-driven rotary presses with automatic feeding devices produce up to 4,000 imprints in an hour. It is a feature of this process that large quantities of ink are needed to produce an image, as well as an absorbent paper which is usually the note paper or at least offset paper type. The copies are better in appearance and permanence than those produced by spirit duplicating, but their cost is somewhat higher.

Offset Lithography

Offset duplicating, as we have said before, is a long-established process in graphic arts. It is based on the transfer of ink from the printing master to the paper through an intermediary elastic surface (rubber roller). Use of this method results in the reduced wear of the master and increased printing speeds. Offset duplicating in conjunction with flat-bed printing is favoured in office reproduction as it offers the possibility of quickly producing relatively large runs several thousand copies with comparative ease of preparation of the printing master.

Office reproduction makes use of simplified small-size printing presses called rotaprints. A rotaprint has three cylinders of equal diameter: the upper is the master cylinder on which the master is fastened, below is the blanket cylinder which is surfaced with a rubberized fabric, and the impression cylinder. Immediately adjacent to the master cylinder are wetting and inking devices which have regulatable supply of the water and ink being fed in. Paper from the feeding tray is automatically supplied into the machine between the blanket and impression cylinders. The impression cylinder presses the paper against the rubberized surface of the blanket cylinder and the ink is transferred from the rubber on to the paper. The imprint thus produced is automatically stacked on the receiving tray. The printing rate can be adjusted within the range of 1,000 to 6,000 copies per hour. The working principle of the rotaprint is shown schematically in Fig. 39 (bottom).

To prepare offset masters, aluminium foil is used which admits of repeated use. However, after the first run the foil must be subjected to graining and complex chemical treatment. When printing small runs, masters made from water absorbent paper are used, which are much cheaper and easier to prepare. Particularly widespread are the electrographic methods of offset masters production.

Offset duplicating can be included with the most expensive and complex processes. The price of a rotaprint greatly exceeds that of other duplicating equipment, and its servicing requires skilled operators. Offset paper must be moisture and dusting resistant, durable, and white, and offset inks must provide increased brightness. But at the same time it is the most perfect of all document duplication methods used in office work. It finds application where there is a steady demand for high quality multiple copies.

Duplicating by addressing machines

Addressing machines (addressographs, adremas) are used primarily for printing the names and addresses of subscribers to periodical publications. They permit repeated reproduction in a specified sequence of a certain set of data. They can be used, however, for the duplication of some information materials as well.

Addressing machines may be based either on the spirit or stencil duplicating techniques, but the predominant method used is printing by embossing (a variety of non-composable letterpress). The masters are prepared by indenting characters in a plastic, fibre-board or metal plate. The laterally reversed raised text thus produced is imprinted on paper through an inked ribbon.

The advantages of duplications using addressing machines are the long life of the imprints, possibility of using any kind of paper, practically limitless runs. But the primary application of these machines determines their deficiencies which are the small size of the master, which contains but a few lines of text, inability to reproduce illustrative materials, low printing speed, labour-consuming preparation of the master. All the above properties of the addressing machines recommend them for use in the libraries for duplicating catalogue cards.

We have thus considered the major techniques and facilities for copying and duplicating of documents. The comparative table of their advantages and shortcomings is given in Fig. 36. As regards selecting particular means of duplication you can be guided in their choice by the table given in Fig. 40. We have made it clear at the beginning of this lecture, that we proposed to set forth only the essentials of reprography, the technical science which deals with the theoretical and methodological aspects of document reproduction processes. If as a result of this discussion you have formed some idea of the new possibilities opening in this field, we shall think that we have fulfilled our task. A knack in handling reprographic equipment comes through special prolonged training. But even this broad overview may help you in making more efficient use of scientific information, and this, after all, is the ultimate purpose of our course. That is why the next (and last) lecture will be devoted to the utilization of scientific information.

Checkup questions

1. What are the basic document copying processes?
2. What are the relative merits and demerits of the different copying methods?
3. What is the distinction between graphic arts and office reproduction techniques?
4. What methods of duplicating are used in office work?
5. What are the criteria for choosing the different methods?
6. What is the main impact of the development of new document reproduction techniques?

Literature

1. Verry H.R. Document Copying and Reproduction Processes. London, Fountain Press, 1960.
2. Verry H.R. Microcopying Methods. London-New York, Focal Press, 1963.
3. Bartlett H.A. A Guide to Reproductive Processes. - "Office Methods and Machines", 1965, vol. 12, 142 p.
4. Schaffert R.M. Electrophotography. London-New York, The Focal Press, 1965.
5. The Revolution in Office Copying. Special Report.- "Chemical and Engineering News", 1964, vol. 42, No. 28-29.

10. UTILIZATION OF SCIENTIFIC INFORMATION

We shall devote this closing lecture to a most difficult - and the least explored - problem of informatics, namely the utilization of scientific information, which is the ultimate goal of all scientific-information activities. This problem in a way brings to a focus all the foregoing discussion since, after all, the diverse primary and secondary documents, the most complex information retrieval techniques and the media and facilities used for it all serve the common purpose of making the utilization of scientific information as efficient as possible. Accordingly, we shall have to pick up afresh some of the points raised in the opening lectures so as to identify the who, when, how and which of scientific information utilization, as well as the modes of scientific-information activities and their developmental prospects.

The role of scientific-information activities

Scientific-information work has an ever-growing role to play in science. It serves essentially to raise the efficiency of science, which is ever more obviously acting as a direct productive force in the present-day society, accounting, as it does, for 3 per cent of GNP of the advanced countries and close on 3 million employees all over the world. The number of scientists and other specialists engaged in scientific-information activities and the sums invested in it have been rapidly growing as well. In the Soviet Union alone, more than 100,000 persons, directly or indirectly, participate in those activities. Over $5,000 million are annually spent on information work in the world. Hence the importance of ensuring the effectiveness of that work in all its aspects: practical, organizational, and theoretical.

We define scientific-information work as an institutionally organized variety of scientific labour performed with the objective of enhancing the efficiency of research and development proper and consisting in the gathering, analytico-synthetic processing, storage and retrieval of recorded scientific information as well as in supplying this information to research and practical workers at a time and in a form convenient to them. Thus far we have been dealing with the processes enumerated in the opening part of this definition. But the process of utilization mentioned in its second part, too, deserves as much attention.

It is no exaggeration that the biggest stumbling-block to an enhancing of the efficiency of scientific information work is the ignorance on the part of the scientist and practician of the opportunities it offers him or his maladroitness in using these opportunities. Science itself as well as the system for dissemination of scientific knowledge have been taking on a mass-scale, one could say, industrial character, but the mentality of the user of scientific information still retains the homespun imprint of the past epoch when the searching and initial processing of information were not a social activity but personal affair of the scientist.

A convincing analysis of this contradiction has been done by M. Menou, a French information scientist. "The most important characteristic of our times, in so far as knowledge transfer is concerned, does not seem to us to be the increase in the number and the power of human knowledge but rather the ratio and the hiatus between our mental structures, so deeply marked by millenaries of "atechnical" behaviour, and the extent of the objects that they must enable us to handle.

Even if they do not always have distinct special characteristics, one may find manifestations and consequences of these prehistoric structures both in scientific and technical activity and in daily behaviour, since being a scientist does not mean that one ceases to be a man".(1)

Principal groups of scientific-information users

In order to learn to get over this psychological barrier, one must get a clear, if only general, notion of the main groups of scientific information users and of the diversity of their information needs. Three user groups are distinguishable according to the kind of activity they are engaged in:

(a) scientists (research workers) mainly engaged in basic and experimental research in the fundamental sciences;

(b) engineers (practical specialists) engaged in experimental design, projecting, and operational activities in technology and other applied sciences or in the industries;

(c) managers engaged in the spheres of science, technology and the national economy.

There is no clearcut borderline between science and technology. Nevertheless, there does exist an essential difference between these two social activities that determines the specific information needs of scientists and of engineers. Science is an activity aimed at production of knowledge and having for its object the cognition of nature. The product of science is knowledge, or scientific information recorded in scientific documents. Technology is an activity aimed at production of useful objects, material wealth. Its task is direct transformation of Nature on the basis of knowledge gained by Science.

D. Price wrote about that difference: "If when a man labours the main outcome of his research is knowledge, something that has to be published open for a claim to be made, then he has done science. If, on the other hand the product of his labor is primarily a thing,a chemical, a process, something to be bought and sold, then he has done technology".(2) A researcher strives at

---

1. Menou M. Knowledge Transfer from Autoconsumption to Mass Production.- In: On Theoretical Problems of Informatics. FID, 435. Moscow, 1969, p.44.

2. Price D. The Difference between Science and Technology. Edison birthday lecture. Dallas, 10th February 1968, p. 11.

learning the unknown, at discovering an order in, of a regular relation between, what seems a chaos of facts. Could he know exactly what knowledge he needs to solve the problem facing him, he would have had much of its solution already in his hand. A scientist thus can never entirely rely on an information service for getting all the information he needs. The greatest task he can entrust to an information service is to select for him the scientific documents that are likely to be of relevance to the subject of his concern.

As a rule, a scientist prefers to scan or peruse those documents himself. In the process he gains a better understanding of his own information needs, comes across valuable associations and guesses, and frames hypotheses. For that reason, of all information services discussed above, most preferable for scientists appear to be: current-awareness information, abstract services, selective dissemination, and retrospective searching.

The case is entirely different as regards the information needs of engineers. An engineer usually knows more or less exactly what kind of knowledge he needs to solve his problem. He is able to state clearly his information needs and consequently expects a clear and concrete answer of the information service. Reluctant to spend time on browsing, he is eager to charge this work on the *information service*.

Information service to engineers thus requires, first and foremost, reference books, analytical reviews, and special services eliciting from scientific-technical literature and unpublished documents information meeting their needs. It was this circumstance that called into being, over and above conventional information units operating in technology and other applied fields, specialized services called "data banks" or "information analysis centres". Over a hundred such services are at present operational in the United States alone.

Information service to managers is again an altogether different matter. The problem is complicated by the fact that the notion of management itself is too broad. We are concerned not with the processes of management and control in general, for example as taken in the cybernetic sense, but only with the particular processes of managing social systems, and primarily science, technology, industry, and economy. In that context management can be viewed as the process of making decisions on the basis of certain available information, or as motivated choice of a course of action out of a range of alternatives. True, any purposive human activity may be viewed as a decision-making sequence, but organizational decisions have their important peculiar features.

A manager usually has to adopt his decisions under the conditions of an acute shortage of necessary information and time to process it, which results in the decisions being to some extent arbitrary. The higher the level of management, the greater this information and time deficits and the more creative management is. If a manager had absolutely exhaustive information and enough time to consider it, the decision-making would not be a creative act, but merely a control function that could be in principle charged to a machine. Of course, we are speaking not of an absolute, but of relative information shortage, since in actual fact a manager is usually flooded by very diverse data. But what he needs is not "crude" information about the object of management but the specific quintessence of that information. And he has no time to extract himself the information he needs for management.

"The management of society and its sections, the organization of human labour is primarily an economic, social and ideological problem, and not a cybernetic, natural-scientific, technical or any other problem."(1) This implies that apart from information about the scientific and technological accomplishments, a manager needs ideological, sociopolitical, economic, and organizational information. Special place in management belongs to prognostic information, since timely and adequate decisions are impossible without an analysis of the developmental trends of the phenomena concerned.

The above considerations suggest that management should have an information service of its own, of which the system of scientific and technical information would make but one component part. Information service to management ought to cover much more sources of information, subject them to a stricter selection, and deeper analytico-synthetic processing, and, moreover, to bring it out faster and reduced to optimal volume. Redundant information hampers decision-making, as the number of alternative strategies tends to grow with increased information volume. However, even a most perfect information service will not relieve the manager from certain information processing that he will have still to perform personally.

These circumstances were pointed to back in 1963 in the report "Science, government and information" submitted to President John F. Kennedy by his Science Advisory Committee:

> "Because information about what is going on is necessary for making management decisions, improvement in scientific information systems is sometimes represented as a panacea for bad management of research and development. Though it is true that poor management can and does occur with the best of communication systems, poor communication almost always leads to bad management."(2)

Study of information needs

The three user groups identified above are very broadly defined. Moreover, they are by no means exhaustive. They do not include, for example, such large and peculiar user groups as teachers and students. A great number of scientists are engaged not in fundamental, but in applied sciences. The information needs of engineers vary widely depending on whether they are concerned with design or

---

1. Afanas'ev, V.G. Nauchnoe upravlenie obshchestvom. (Opyt sistemnogo issledovaniya). (Scientific management of society. An essay of system research). Moscow, "Politizdat" Publishers, 1968, p. 379.

2. Science, Government, and Information. The responsibilities of the Technical Community and the Government in the Transfer of Information. A report of the President's Science Advisory Committee, Washington, US Government Printing Office, 1963, p.10.

operation work. Many scientists and engineers are at the same time managers. Information service in many fields of science and practical activity, in particular in the social sciences, medicine, agriculture, has certain specific features. All these and many other aspects of scientific information utilization require a systematic study.

Information user needs are studied by sociological research methods, with ever deeper exploration of the psychology of scientific creativity. This problem, which surpasses the scope of informatics, has not yet been adequately investigated, although hundreds of studies are devoted to it every year in different countries. Only a minority of these are of a general nature and strive at establishing the common regularities of scientific information utilization, while most pursue for narrower practical ends: to improve information service to this or that particular user group, or establish the optimum pattern of functioning of particular information agencies or networks. As yet this research is conducted as scattered efforts and lacks any common methodology, with the result that the findings are difficult to compare. Still, some common objects and methods of study are already identifiable.

First of all, information users themselves are explored, their typology, dependency of information needs on their vocational pursuits, purposes of reference to information, qualifications, age, professional rank and a range of other characteristics. Some findings suggest that the occupation of the user and the kind of establishment he is employed by are the factors determining his information needs. The different working conditions influence the time spent by the user on information retrieval, utilization of sources, and frequency of reference to libraries and information agencies.

It depends on the field of activity whether the organizational structure of user groups is worth studying. Big, institutionalized user collectives are studied mostly in technology and applied sciences as well as in such natural sciences as chemistry and physics. In biology and other life science, information utilization in small research teams; and in social sciences and the humanities, individual user needs; are in the focus of attention.

In this connection it is important to explore the activities of the different types of information agency, the modes of information service, and channels. No less attention should be paid to the study of information sources, the dependence of the utilization of the various primary and secondary documents on the kind of user activity and other user characteristics.

All these and other aspects of information utilization are explored by sociological methods. Questionnaires are broadly applied. They are distributed among a representative sample of the user group studied. The replies are statistically processed and interpreted, so that the findings either corroborate or refute a previously formulated hypothesis about this or that relationship of factors influencing scientific information utilization. The method is usually supplemented by interviews with those users whose replies are unclear or hard to interpret. Interviews can also be applied in the case when it would be difficult to formulate questionnaire questions, provided the group studied is not too large.

Other methods worth mentioning are: observation, and keeping diaries by the users themselves. Complex studies using system engineering and operations research techniques have been carried out of late. Promising results have been achieved in some aspects of information utilization studies by analysis of citations in journal articles, feedback from SDI system users, and analysis of requests coming to libraries and scientific information agencies.

All these methods, however, have their limits since they are based on user judgments that are naturally biased. The constant contact of a user with his habitual information system results in his no longer expecting of the system anything surpassing its capacities. This habit becomes a stable psychological factor likely to substantially distort the results of a study based on some of the above methods or their combination. After all, the purpose of a user study is to learn their actual needs and not what they think are their needs. This suggests that for studying the different types of information needs at the various stages of research and design work it is necessary for the explorer personally to participate in that work.

Types of information needs

The most general types of information needs have been defined by M. Voigt, an American information scientist, in the early 1960's.(1) By questioning a hundred Scandinavian scientists in chemistry, physics, and biology he ascertained that they refer to information sources mainly in the three cases:

(1) while getting current awareness of the results both in their particular narrow field and the related disciplines;

(2) in their day-to-day work, when they need some particular factual information on figures, methods and designs; and

(3) when embarking on a new problem or project as well as when completing it and proceeding to commit it to paper - they need then a retrospective search to identify as many published and unpublished sources as possible on the subject concerned.

The first case may be characterized as the natural need of a scientist or other specialist for keeping au courant with the developments in science, technology or culture; the second case is the everyday utilization of information for reference purposes; and the third case is an exhaustive approach to scientific documents relating to a particular topic or subject.

These types of user needs have been since then repeatedly re-interpreted and specified in keeping with the different aspects of user studies. For example, as regards technology it has been suggested that three stages at which information needs arise should be considered: the choice of the trend of an R & D or design project, its planning, and the execution of the project. In choosing the

---

1. Voigt, M. Scientists' approaches to information. New York, ALA, 1961.

trend the user needs general information on the existing approaches to the problem of his concern, on their relative advantages and drawbacks both of essential and of organizational-technical nature. When planning a project in conformity with the trend chosen he needs more detailed data on the resources and their distribution (time, money, personnel, equipment, room). At last, at the execution stage, what he needs is very detailed information on all the elements of the object under development.(1)

A more generalized approach to the differences in the information supplied at the various stages of scientific and technical work is developed in the collective monograph by a group of Ukrainian scientists entitled "Information in scientific research" and published under the editorship of Ukrainian Academician A.N. Sccherban' (1) They distinguish as many as seven such stages:

1. Overall familiarization with the problem and problem statement. Drawing up a plan and the provisional terms for the solution of the problems of primary and of secondary importance. This stage comprises general acquaintance with information available on the subject.

2. Gathering scientific knowledge about the subject of study. Retrospective searching of the broadest possible scope of literature without any pronounced critical approach.

3. Coordination and interpretation of scientific data. A critical evaluation of the ideas and hypotheses of the different authors. The relevance criteria for the information needed are specified and the amount of documents used is reduced.

4. Statement of hypotheses and choice of the working hypothesis. This is the most important stage of research in technical sciences. As to information utilization, it is characterized by deep processing rather than broad coverage. Still, sometimes the working hypothesis may be chosen under an influence of what might appear "irrelevant" information.

5. Proving the working hypothesis. The most important stage in basic research. Information used depends on the specifics of research, for instance, the proving of an assumption may require gathering factual data scattered in the literature.

6. Statement of conclusions and recommendations. Alongside with generalizations, predictions are frequently made at this stage. Information is often used to shed light on precedence and priority aspects.

---

1. Bernshtein, E.S. Of informatsionnykh potrebnostyakh i kachestvennom preobrazovanii inforamtsii. (On information needs and qualitative transformation of information).- "Nauchno-tekhnicheskaya Informatsiya", 1967, ser.2, N.6, p.8.

7. Description of the research results. Information gathered and processed is, as a rule, minutely documented.(1)

The above schemes are, of course, very far from perfect, and they may be greatly modified depending on the peculiar features of this or that study. However, even as such they are, they are evidence of the close association between the utilization of scientific information and the logical structure of scientific research.

There have been attempts, too, to asses the utilization of the different sources as depending on the kind of activity pursued by the user and the stages of scientific research process. It has been noticed, for example, that abstract journals, bibliographies and other secondary documents are more intensively used at the outset and at the end of elaborating a problem, while the need for reference aids and primary documents (particularly special types of technical literature) is more acutely felt at the stage of experimental testing of the theoretical assumptions, equipment, process or product specimens.

A great many studies concerned with the channels through which scientists receive information testify to an "underemployment" of information services and publications. While the numerical results of these studies which use different methods and different user groups vary greatly, it is almost regular that among the most important sources yielding more than a half of all information needed for research such "accidental", not purposive and for the most part oral sources are named as talks with colleagues, and conference reports. Much less information (10-20 per cent) is extract from the abstract journals and bibliographic indexes, and even less from library catalogues. Thus, for all the inadequacies of methodology and unreliability of the findings of user studies, it may be asserted as an indisputable fact that the efficiency of scientific information activities still remains very low. The problem of evaluating this efficiency arises in this connection.

Efficiency of scientific information

This problem has become a topical one since scientists and other specialists feel themselves constrained to spend ever greater share of their time on searching for information instead of on research proper. A recent study of the time balance of American chemists showed that they spend an average of 45 per cent of their working time looking for scientific and business information, while all their experimental work accounted but for 36 per cent, and the processing of the results obtained, less than 6 per cent.

---

2. Informatsiya v nauchnykh issledovaniyakh. (Information in scientific research) Otv. red. akad. AN USSR A. N. Sccherban'. Kiev, "Naukova Dumka". publishers, p. 119-120.
(Nauka, uchenyi, informatisiya, /Science, scientist, information/ by S.G. Zlochavskii, A.V. Kozenko, V.V. Kosolapov, A.N. Polovinchik).

The efficiency of scientific information activity is a comparatively recent concept. It still lacks a unique definition and is diversely interpreted, now as the total revenues from the activity, now as the surplus of revenue over the expenditure, now again as the ratio of revenue to expenditure. This "economic" approach to scientific information activities stems from the fact that ever greater funds are invested in it and the requirements placed on it are growing. "The last factor gives rise to two questions: what benefits can be derived from information activities by institutions and enterprises which organized the information service or by the national economy (in case of socialist countries where national information systems are established), and what is the correlation between profits obtained and expenses incurred?".(1)

It is relatively easy to count up the expenditure on scientific information activities, which is, incidentally, quite sizable as we have seen. As to the economic benefits, they are much harder to gauge, since as often as not they lack any direct economic expression. The difficulty is similar to that of numerical evaluation of the effect of education. Therefore, the existing system of the evaluation of economic activities proves inapplicable in this case. The same Polish information scientist whose words were quoted in the preceding paragraph suggests that information efficiency should be measured by the ratio of the amount of relevant information supplied to the users to the total amount of information gathered and processed at the information centre concerned.(2) He has formulated the corresponding efficiency criteria and calculation techniques and relationships.

However, other difficulties arise, too. We have already seen how relevance judgments may be subjective in measuring information system efficiency. In this case as well, the assessment criteria are mostly based on a correspondence between the contents of the documents gathered, on the one hand, and potential user needs and actual user requests on the other. But we are yet unable to measure the objective value of scientific information in the same manner as the mathematical information theory measures information contained in a message by the uncertainty this message serves to eliminate. And, it is not the knowledge of some individual about the object or phenomenon described in the message that the measure of this uncertainty depends on but only the objective properties of that object or phenomenon.

Despite the fact that scientific cognition is social by nature, we are able to measure the value of scientific information only with reference to separate individuals or their groups. Human society does possess a social mechanism which objectively evaluates scientific information, but unless and until this mechanism is explored and understood the measurements of the efficiency of scientific information activity are bound to remain empiric.

---

1. Piróg, W. Efficiency of Information Activities: Evaluation Criteria, Methods and Indices. - In: On Theoretical Problems on Information. FID 435. Moscow, 1969, p. 93.

2. Ibid, p. 94.

The contents of scientific information activities

We have ascertained that any creative work may be viewed as the process of selection and processing of information for different purposes: in scientific research - to create new scientific information; in engineering and industrial activities - to build new machines, and develop new substances and engineering procedures; and in the management sphere for taking organizational decisions. But, scientific information activity itself to a great extent consists in selection and processing of scientific information. What then are its specific features and what is the borderline between scientific information work and other creative activities?

Two diametrically opposed replies have been given to that question: the first viewpoint debases scientific information work just to a variety of auxiliary service to research, design, and science-management activities; the other viewpoint fully identifies it with the rest of research labour. Neither point of view appears tenable. If the scope of scientific information work is limited to a variety of ancillary or technical assistance rendered to scientists and other specialists in furnishing them with documents they want, why has it become established as an activity in its own right in the face of library and bibliographic activities that supposedly discharge the same function? This view engenders a dangerous illusion that the "information crisis" could be cured simply by better special libraries without any essentially new methods and media of scientific communication. This point of view virtually negates scientific information work as an independent activity, but this is contradicted by the actual social practice.

On the other hand, broadening the scope of scientific information activities so that they encompass genuine research as well, is no less wrong; eventually this, too, boils down to a denial of its independent status. The erroneous identification of basic research with scientific information work stems from the fact that both the theoretical scientist and the informationist are engaged in analysis and synthesis of scientific information, at the stages of its selection and especially processing and that they both deal almost exclusively with scientific documents. By outward appearance, the work of U. Le Versier who computed the Orbit of the Neptune, a planet then unknown, or of D.I. Mendeleyev who discovered his periodic law resembles the work of an information officer who prepares for an explorer the information he needs, extracting it from scientific documents and presenting it in a form convenient to the explorer. But, for all the outward likeness, the two kinds of work have evidently differing goals. While in the former case the goal is to discover scientific laws and create new scientific information, in the latter it is only to provide prerequisites to authentic research.

Naturally, the border between the two kinds of activities is conventional and shifty: it depends on a great number of factors, in particular the stage of development of the science in question, the traditions existing in it, the level of formalization it has reached, and a number of others. Still, this borderline is always clearly discernible at any moment in any particular conditions. Scientific information work constitutes, as it were, the opening and preparatory phase of any research. It creates premises for a more efficient investigation by selecting the documents pertinent to the subject of research and providing the scientist or expert with the information he needs, extracted from those documents and tailored to suit his wants. Part of that work is usually performed by the user himself. It is in this sense that it is sometimes said that the researcher, too, carries out scientific-information work, which is not exact since the term should be applied only to the socially and institutionally organized forms of activity.

Scientific information activities as well as the system of scientific publications in which these activities first emerged existed and still exist as a natural condition of the development of science. Since their very emergence they were unfolding within science as an important means of scientific communication. For ages, scientists, supported only by libraries, themselves searched, selected and processed scientific information. This was the reason that some important forms and methods of scientific information work first took shape in the library - until recently the only social institution to store, analyse and generalize the age-long experience in the field.

The advancement of science in the last decade, and its turning into a major and mass-scale human activity, as well as the gradual penetration of science by the industrial methods and organization of management all led to establish scientific information work as an organic but autonomous part of scientific labour. We have spoken in the first lecture about the similarity of this separation to the division of research into theoretical and experimental. We can now bring more precision to this.

The process of social division of labour in present-day science is unfolding at least along two different lines. Research divides into basic and experimental primarily according to the methods employed. Simultaneously, a similar process is developing along a different, notably the functional line, where three independent varieties of scientific work are clearly discernible - research proper, scientific information activities, and science management.

Outlooks

When discussing the contents and place of scientific information work in science it would be wrong to pass over its future prospects. This is the more essential as the scope of scientific information activities is continually expanding. C.W. Mooers, a pioneer of automated information retrieval, said in 1959, in a report at a conference: ... "Information retrieval" is a term that I had the pleasure of coining only ten years ago, at another computer conference. The term has come a long way since.

In thinking about a definition of information retrieval, and in considering the future of this field, we must take an evolving view. At the present time, information retrieval is concerned with more than the mere finding and providing of documents; we are already concerned with the discovery and provision of information quite apart from its documentary form. To encompass future developments, as we shall see, even this broad view of information retrieval will have to be modified and extended".(1)

A half of the term of this prediction now being over, one is entitled to say that it has proved to be valid. The proposition is especially important in forecasting information activities development which has its own regularities and cannot be constructed in an arbitrary manner.

The regularities of the growth of science itself and the peculiar traits of the present stage of scientific and technological revolution are doubtless the key factors affecting the growth of scientific information activities.

There are many facts in the practice of information service to indicate a trend towards a further shift of tasks within the scope of research proper to the sphere of scientific information work. The common eagerness of the users to get as much information as possible spending the minimal effort is bound ever more insistently to press on the information systems the tasks of synthetizing, creatively processing and tailoring information to user needs. Continuing interaction between information systems and the different user categories will ultimately enroll users' participation in system optimization and heighten the contribution by scientific information work to research endeavour.

To summarize: in this last lecture we profiled the role of scientific information activities under a new angle, discussed the varying patterns of scientific information utilization by the main user groups, highlighted the principal objects and techniques of the user needs studies and the main types of user needs as arising at the different stages of research and design work, touched upon efficiency evaluation problems, specified the contents of scientific information activities, and outlined its development prospects.

As we have seen, the better organized will be research conducted in the different fields of science and technology and the more formalized will be the special languages and methods used in those fields, the more sophisticaled information selection and processing tasks will be charged to institutionally independent information services. Knowledge of the basic regularities, methods and media of scientific information work is indispensable for coping with these tasks. Providing clues to that essential knowledge is the purpose of these lectures.

---

1. Mooers C. The Next Twenty Years in Information Retrieval. Some Goals and Predictions. - "American Documentation", 1960, vol. 11, N 3, p. 229.

Checkup questions

1. What principal groups of scientific information users are distinguishable? What are the differences between their information needs?
2. What are the objects and techniques of user needs' studies?
3. How is scientific information utilized at the different stages of research and design work?
4. How can the efficiency of scientific information work be evaluated?
5. Where lies the borderline between research proper and scientific information work?
6. What are the major developmental prospects before scientific information activities?

Literature

1. Fairthorne R.A. The Scope and Aims of the Information Science and Technologies.- In: On Theoretical Problems of Informatics. Fid 435. Moscow, 1969. p. 25-31.
2. Wysocki A. Study of User's Information Needs: Subject and Methods. - In: Ibid., p. 80-92.
3. Information Systems Workshop. The Designer's Responsibility and his Methodology. Washington, Spartan Books, 1962. V, 153 p.
4. Kent A., Taulbee O.E. (eds.). Electronic Information Handling. Washington, Spartan Books, 1965. VII, 355 p.
5. Scientific and Technical Communication. A Pressing National Problem and Recommendations for its Solution. A synopsis of the report of the Committee on Scientific and Technical Communication of the National Academy of Sciences - National Academy of Engineering. Washington, 1969. 30 p.

SYLLABUS AND PROGRAMME

for the

INTRODUCTORY COURSE ON INFORMATICS/DOCUMENTATION

1. The subject and methods of informatics (documentation). Laws of scientific development that brought about the establishment of scientific information activity in the status of independent discipline : differentiation and specialization, traditions and international scale of this activity, its accelerated development. Objectives and stages of scientific information activity: collection, analytical and synthetic processing, storage, retrieval, dissemination, and use of scientific information. Fundamental concepts: information, scientific information, scientific information activity, informatics (documentation). Interrelations between informatics and the information theory, cybernetics, semoitics, linguistics, psychology, library science, bibliography, history of books, science of science, and the technology.

2. Scientific literature as a medium of knowledge dissemination (primary scientific documents and publications). Phases of evolution of written documents: cuneiform writing tablets, papyrus scrolls, parchment codices, manuscript paper books, printing, modern methods of publishing. Scientific documents and their kinds: published and unpublished, primary and secondary documents. Description of the principal types of primary document: books (monographs, collective volumes, conference proceedings, textbooks and manuals, official publications, book series; book statistics); periodical publications (journals, serials, newspapers); special technical publications (standards, paten specifications, technical catalogues); unpublished primary publications (scientific-technical reports, theses, preprints, deposited manuscripts). The regularities of distribution and obsolescence of scientific publications (Bradford's law of Scattering, bibliographic coupling of documents, obsolescence of books and of journal articles, the "half-life" of scientific literature).

3. Secondary scientific documents and publications. Conventional secondary documents: reference literature (encyclopaedias, reference handbooks, dictionaries), surveys (e.g. "Progress in Science", "Scientific and Technical State-of-Art Communications"), abstract journals (specifying the goals, scope, volume, and pattern of the principal abstract journals in natural and social sciences: "Chemical Abstracts", "Biological Abstracts", "Science Abstracts", VINITI's "Referativnyi Zhurnal", "Bulletin signaletique", etc.), current awareness bulletins, library catalogues/classified, subject, dictionary, alphabetic/, bibliographic indexes (main types as defined by contents, purpose, date and place of publication: "Cumulative Book Index, "British National Bibliography", "Bibliographie de la France", "Deutsche Nationalbibliographie, "Deutsche Bibliographie", "Knizhnaya letopis SSSR", etc.; survey of special retrospective bibliographies and the bibliographies of bibliographies. (Besterman, Malcles, Totok und Weitzel, etc.); Unconventional indexes: permuted (KWIC),KWOC,tabledex, citation indexes; current awareness publications and services (current contents, selective dissemination of information).

4. Organization of Scientific Information Activities. Scientific information channels: personal contacts and written communication of scientists, exchange of pre- and reprints, participation and scientific meetings; invisible colleges. Editing and printing processes and their place in scientific information activities. Relationships between libraries and information units. Information agencies and facilities: systems of information serving specialists in the leading countries, description of the biggest information agencies, informational practices and procedures. Special (research) libraries: national and largest independent universal libraries, libraries of R & D bodies, university libraries, and libraries in industry; principal library practices and procedures. International organizations engaged in scientific information and library activities: Unesco, FID, IFLA, ISO TC-46, ICSU, and others.

International cooperation in these fields.

5. Fundamentals of information retrieval. Information retrieval, information retrieval systems /document and fact retrieval/, information retrieval languages, retrieval criteria /relevance and pertinence of documents/, performance of retrieval systems and methods for its assessment.

6. Conventional information retrieval systems. Library regarded as an IS & R system, library catalogues. Alphabetical retrieval files, the methods of bibliographic description. Classified files: principles of classification, hierarchical library and bibliographic classifications, UDC and its application. Facet classifications. Subject classifications and subject indexing procedures.

7. Nonconventional information retrieval systems. Coordinate indexing and its distinctions from conventional indexing systems. Descriptor languages and their creation: synonymy and multiple-meaning control. Thesauri: general principles, main types of thesaurus, representation of paradigmatic and syntagmatic relations,methods of thesaurus construction. Specialized information retrieval languages: WRU semantic code, SYNTOL.

8. Hardware of information retrieval. Retrieval devices operating with discrete information media: catalogue file cards, uniterm cards, punched cards (manual and machine-sorted), microfiches ("Filmorex", "Minicard"). Retrieval devices operating on continuous media: punch tapes, roll microfilm (microfilm selectors: "Rapid Selector" system, "Miracode"), magnetic tape ("Videofile"). Computer-based implementation of information retrieval systems.

9. Reproduction of documents. Document copying: silver processes /photography, reflex processes; stabilization, diffusion, matrix Verifax/; diazo processes (diazocopying, Calfax); thermographic processes (Thermofax, Masterfax, Imagic); electrographic processes (xerography, electrofax, transfer, and thermoplastic processes). Document duplication: printing processes (relief and plane printing, photogravure); typesetting machines; offset, stencil, spirit duplication, addressing machines.

10. Utilisation of scientific information. Study of information needs and requests. Efficiency of information utilization. Efficacious methods of document handling. Influence of scientific information activities upon the development of science, technology, economy and culture.

		Hours		
		Lectures	Seminars	Total
1.	The subject and methods of informatics/documentation	2		2
	Source of scientific Information			
2.	Scientific literature as a medium of knowledge dissemination	2	2	4
3.	Secondary scientific documents and publications	4	4	8
4.	Organization of scientific information activities	4	2	6
	Information Retrieval			
5.	Fundamentals of information retrieval	2		2
6.	Conventional information retrieval systems	4	4	8
7.	Nonconventional information retrieval systems	4	2	6
8.	Hardware of information retrieval	2	2	4
	Use of Scientific Information			
9.	Reproduction of documents	2	4	6
10.	Utilization of scientific information	2		2
	Total	28	20	48

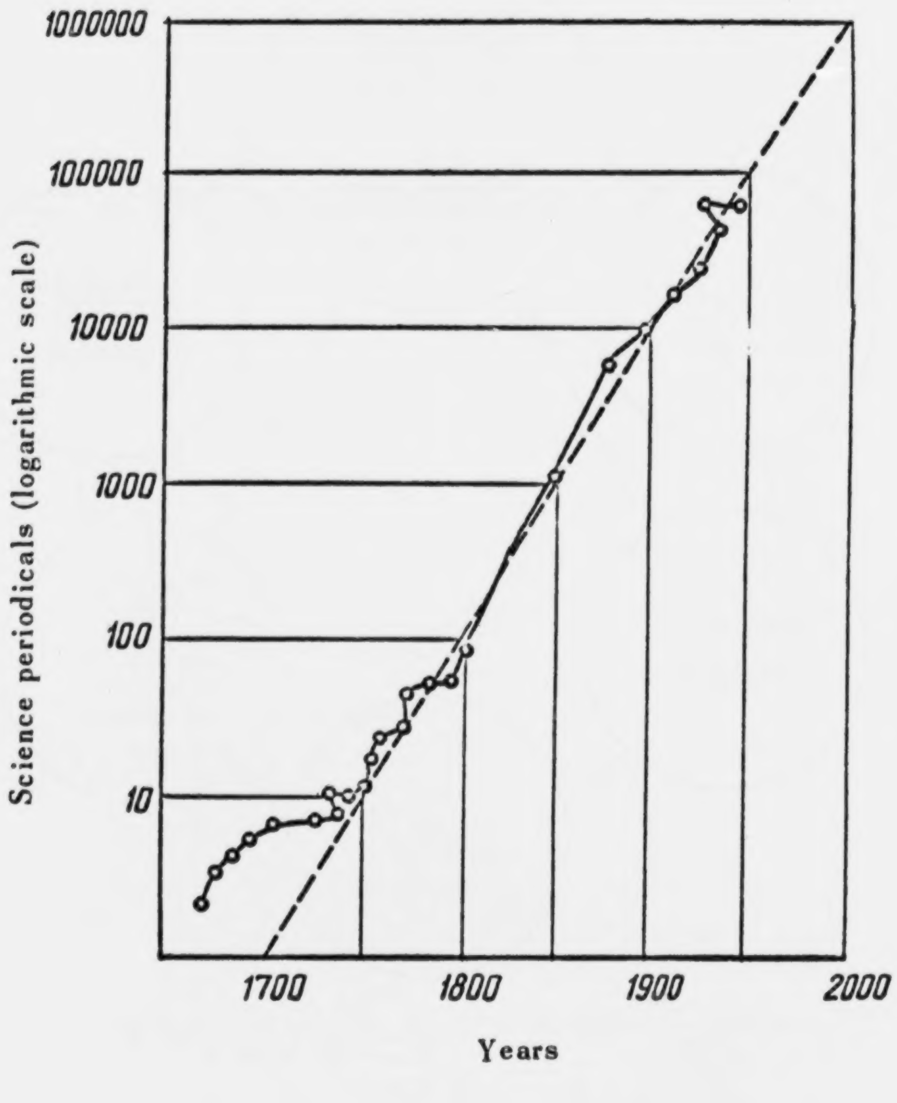


Fig. 1. Growth of world science periodicals

The diagram shows the total number of periodicals, including those no longer published. The notion of a "science periodical" has been taken in a broad sense. Adopted from: Price D.J. de S. Science since Babylon. New Haven, 1961, p. 97.

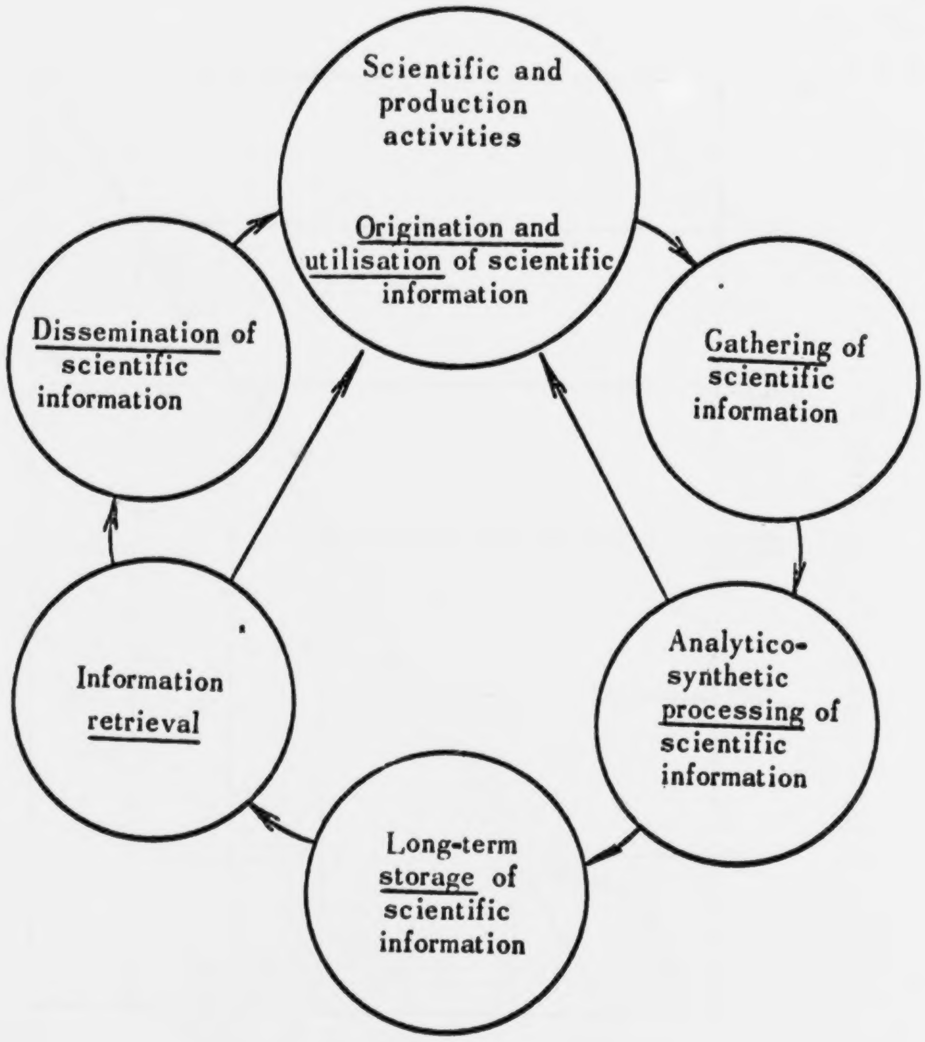

Scientific and production activities
Origination and utilisation of scientific information
Gathering of scientific information
Analytico-synthetic processing of scientific information
Long-term storage of scientific information
Information retrieval
Dissemination of scientific information

Fig. 2. Tasks and stages of scientific-information activities

Origination of scientific information is not included in scientific-information activities by some authors. However, origination as well as utilisation of scientific information make up the organizational sphere of scientific labour and are obviously major aspects of scientific-information activities as a whole.

Scientific documents and publications		
	Primary	Secondary
Published	Books and pamhplets	
	Monographs Collections of papers Conference proceedings Textbooks and manuals Official publications	Reference literature
	Periodical publications	
	Serials Journals and magazines Newspapers	Surveys Abstract journals Express information bulletins
	Special types of technical publications	
	Standards Invention specifications Technical catalogues Information leaflets	Standards indexes Invention bulletins Patent classifiers
Unpublished	Scientific and technical reports Dissertations Information cards Translations Preprints Manuscripts and galleys Data files	Library catalogues Bibliographic files

Fig. 3. Types of scientific documents and publications.

For clarity each type is mentioned only once. Actually many types could be indicated in more than one place, for example, bibliographic indexes are not always periodical, library catalogues may be published, etc.

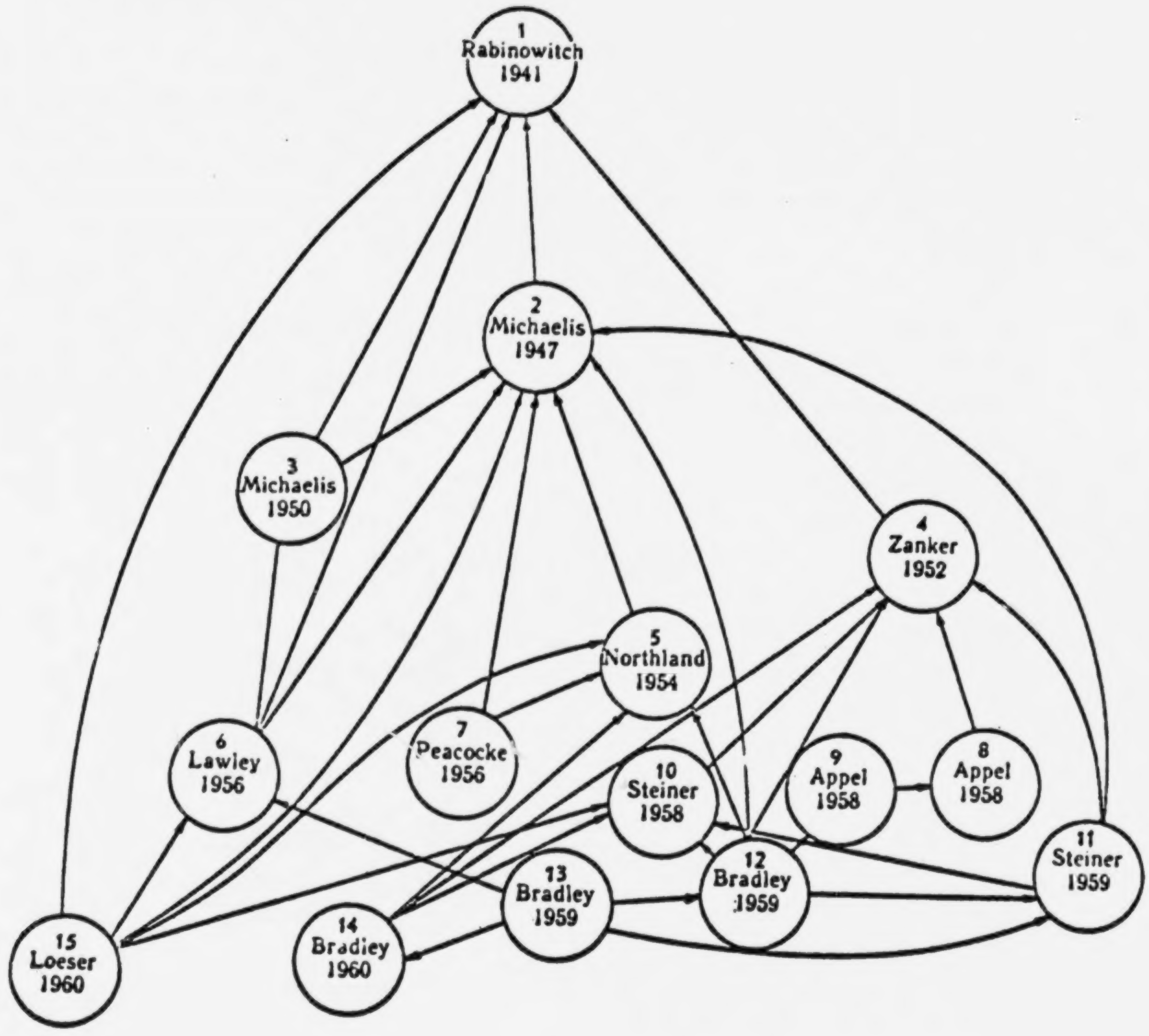

1
Rabinowitch
1941
2
Michaelis
1947
3
Michaelis
1950
4
Zanker
1952
5
Northland
1954
6
Lawley
1956
7
Peacocke
1956
8
Appel
1958
9
Appel
1958
10
Steiner
1958
11
Steiner
1959
12
Bradley
1959
13
Bradley
1959
14
Bradley
1960
15
Loeser
1960

Fig. 4. Citation of science documents in later publications.

Circles indicate works on nucleic acids published between 1941 and 1960, e.g.:

1. Rabinowitch E. and Epstein L.F. Polymerization of Dyestuffs in solution. Thionine and methylene blue. – "Journal of American Chemical Society", 1941, v. 63, № 69.
10. Steiner R.F. and Beers R.F. Spectral changes accompaning binding of Ao. by Poly A. – "Science", 1958, v. 127.

Arrows show references. The diagram is by G. Allen (Carnegie Institution). Adopted from the prospectus to "Science Citation Index" for 1963.

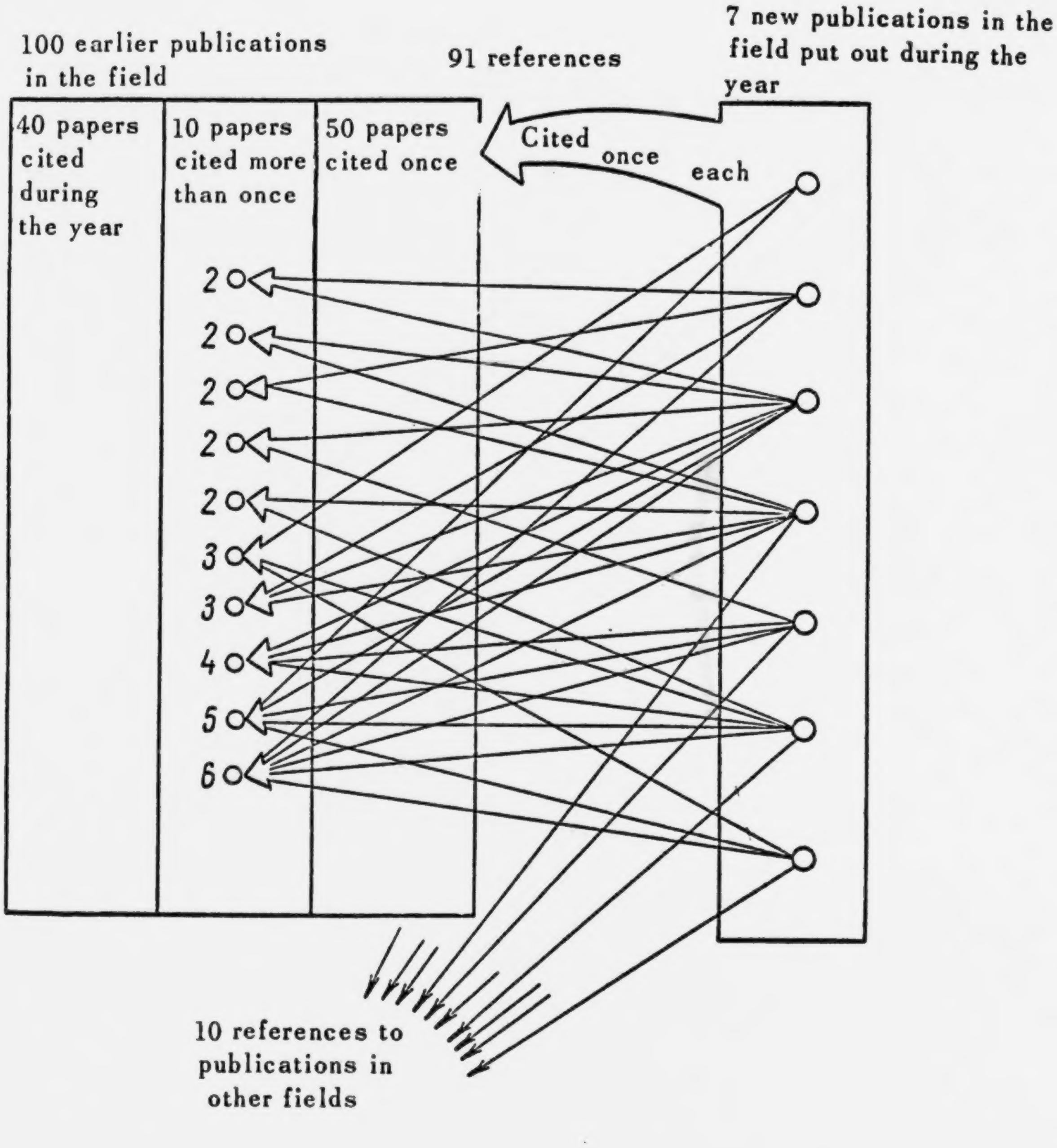

100 earlier publications in the field
91 references
7 new publications in the field put out during the year
40 papers cited during the year
10 papers cited more than once
50 papers cited once
Cited once each
2
2
2
2
2
3
3
4
5
6
10 references to publications in other fields

Fig 5. Distribution of new references to earlier publications in a hypothetical field.

The number of publications was arbitrarily set equal to 100, the rest of the data refer to the average values for the different fields of knowledge derived from an analysis of the 1961 Science Citation Index. To make the diagram more illustrative an alomost "closed" field is shown so that only 10 references are to publications in other fields. Due to the small number of papers in the right-hand column (7), the distribution of references they contain could not be shown adequately in conformity with the regularities indicated above.

Adopted from: Price, D.J. de S. Networks of scientific papers. — "Science", 1965, v. 149, № 3683, p. 512.

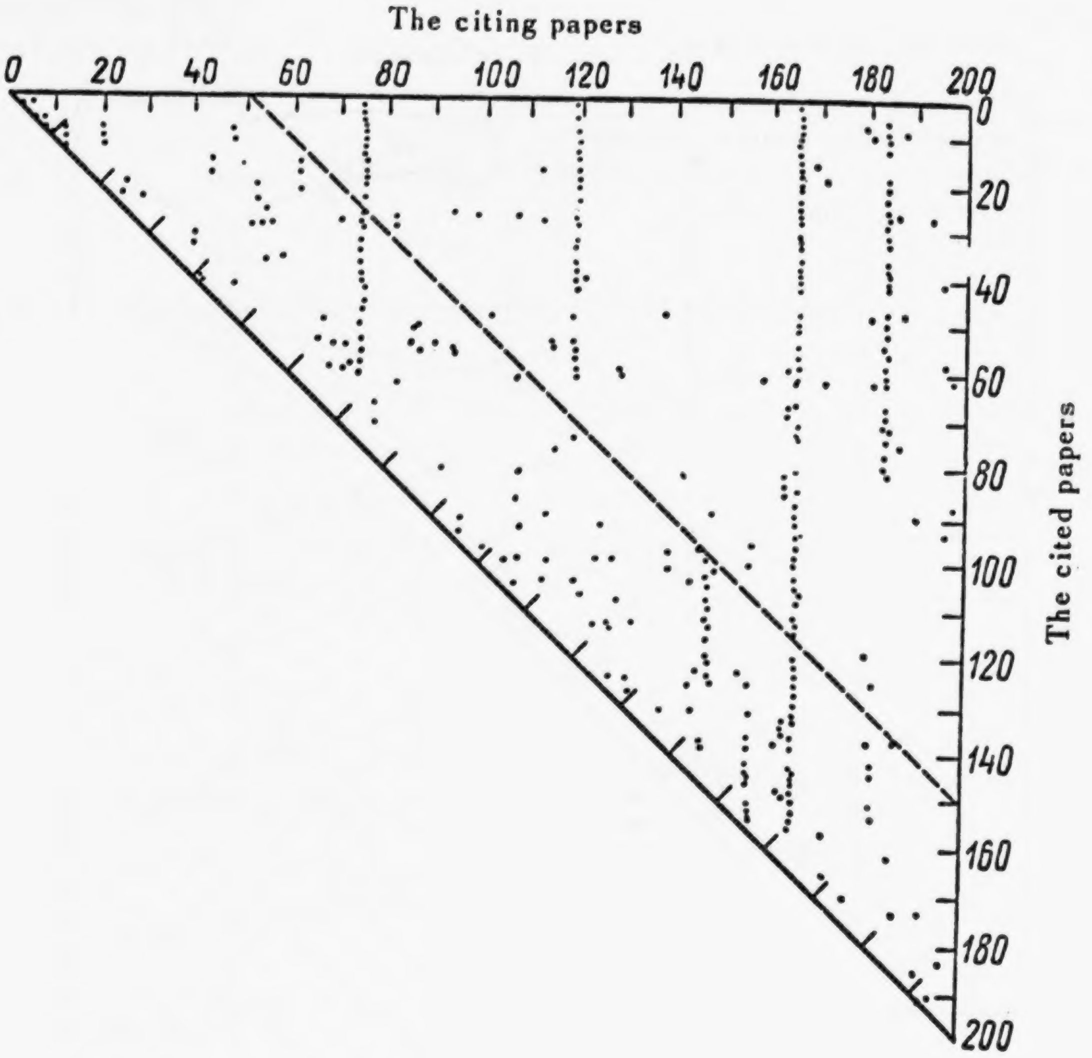


Fig. 6. Scatter of citations in an isolated special literature.

The horizontal and the vertical axes indicate numbers of the same papers ordered chronologically. All the papers refer to a narrow physical subject – the study of imaginary N-rays falsely discovered in 1904. Each column of dots designates the references contained in the paper indicated by the horizontal number opposite that column. Every dot designates a reference to the paper indicated by the number on the vertical axis opposite to it.

Adopted from: Price, D.J. de S. Networks of scientific papers. – "Science", 1965, v. 149, № 3683, p. 514.

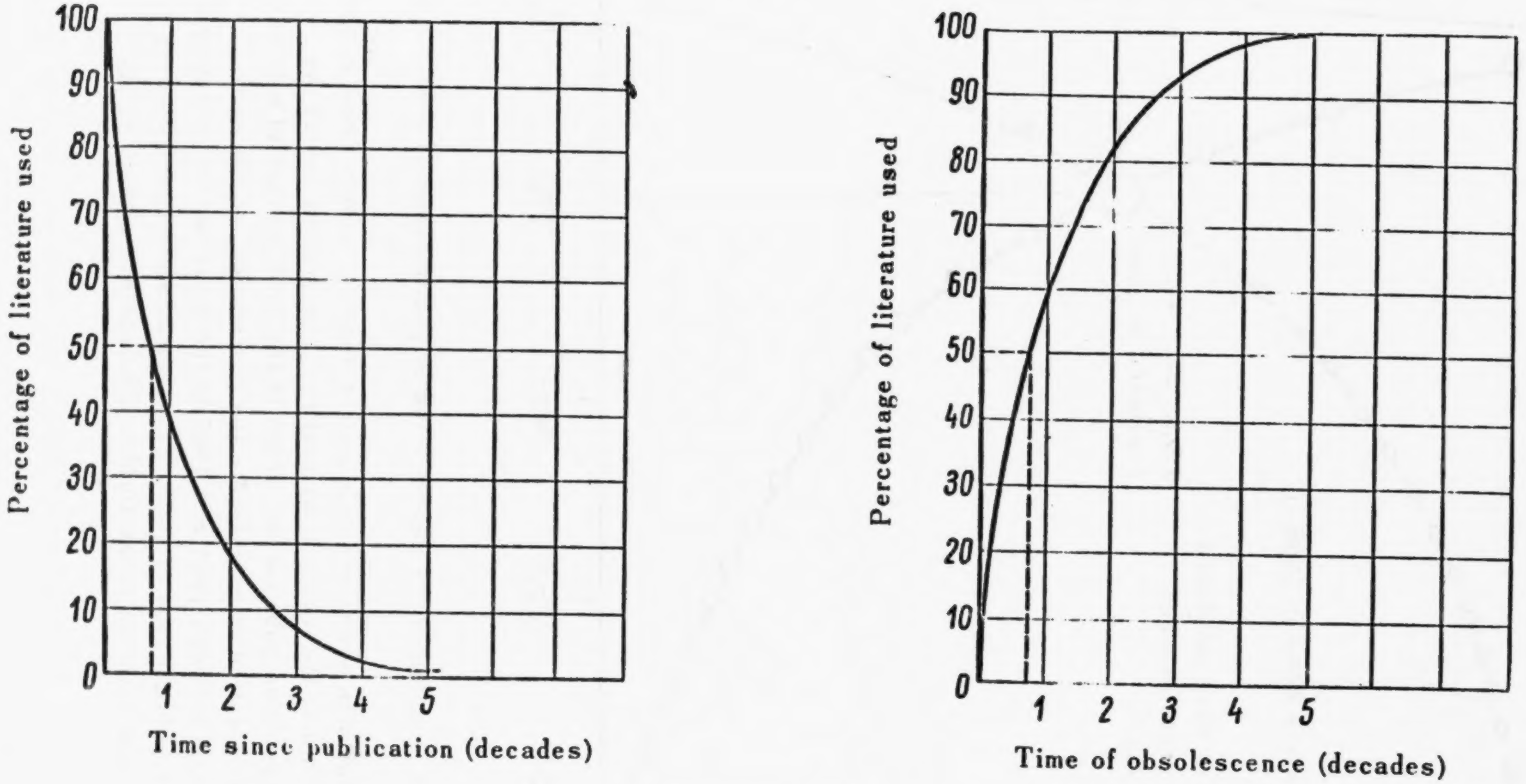


Fig. 7 Half-life of scientific literature.

Adopted from: Burton R.E. and Kebler R.E. The "half-life" of some scientific and technical literatures. – "American Documentation", 1960, vol. 11, № 1, p. 18–20.

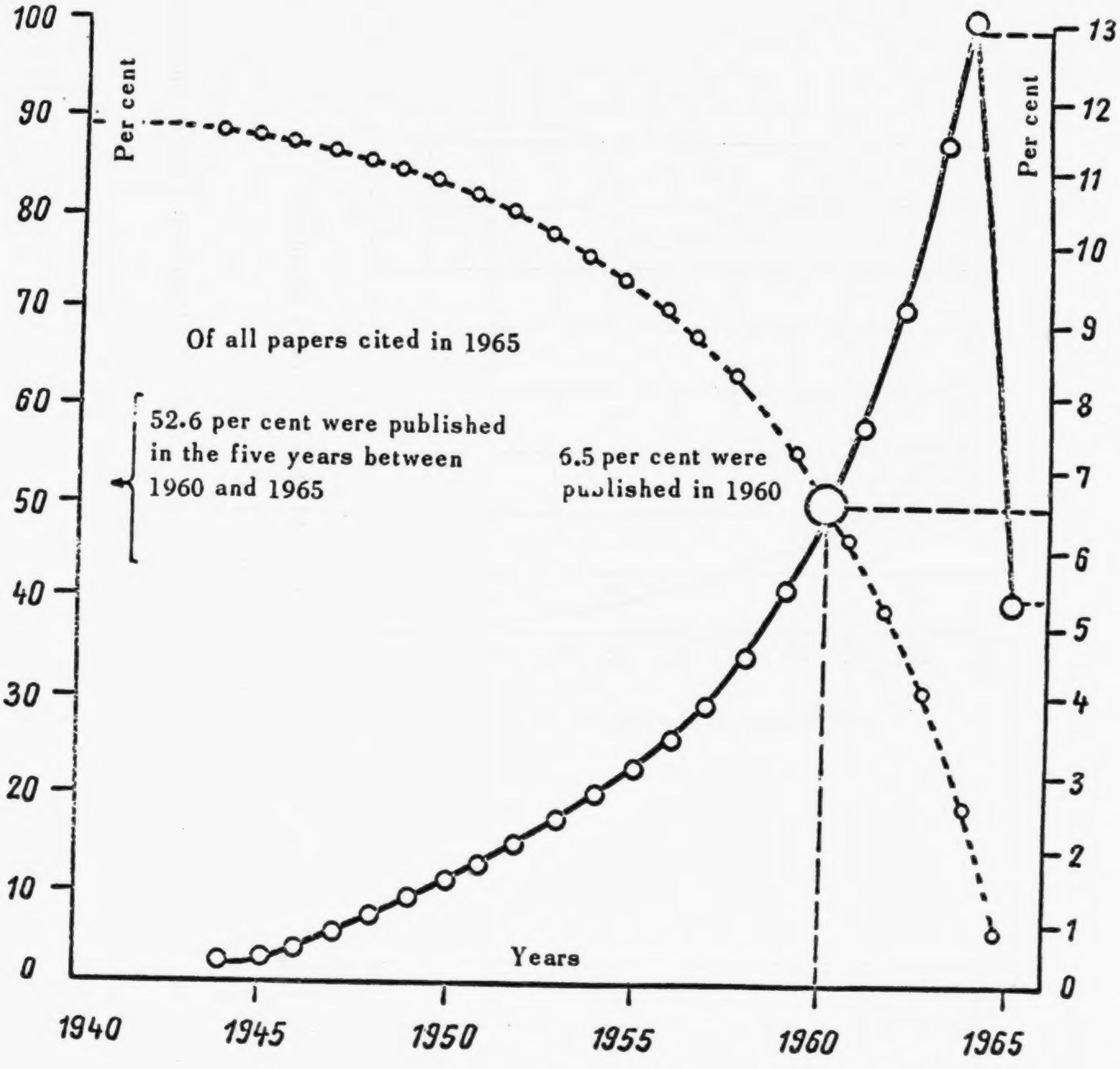
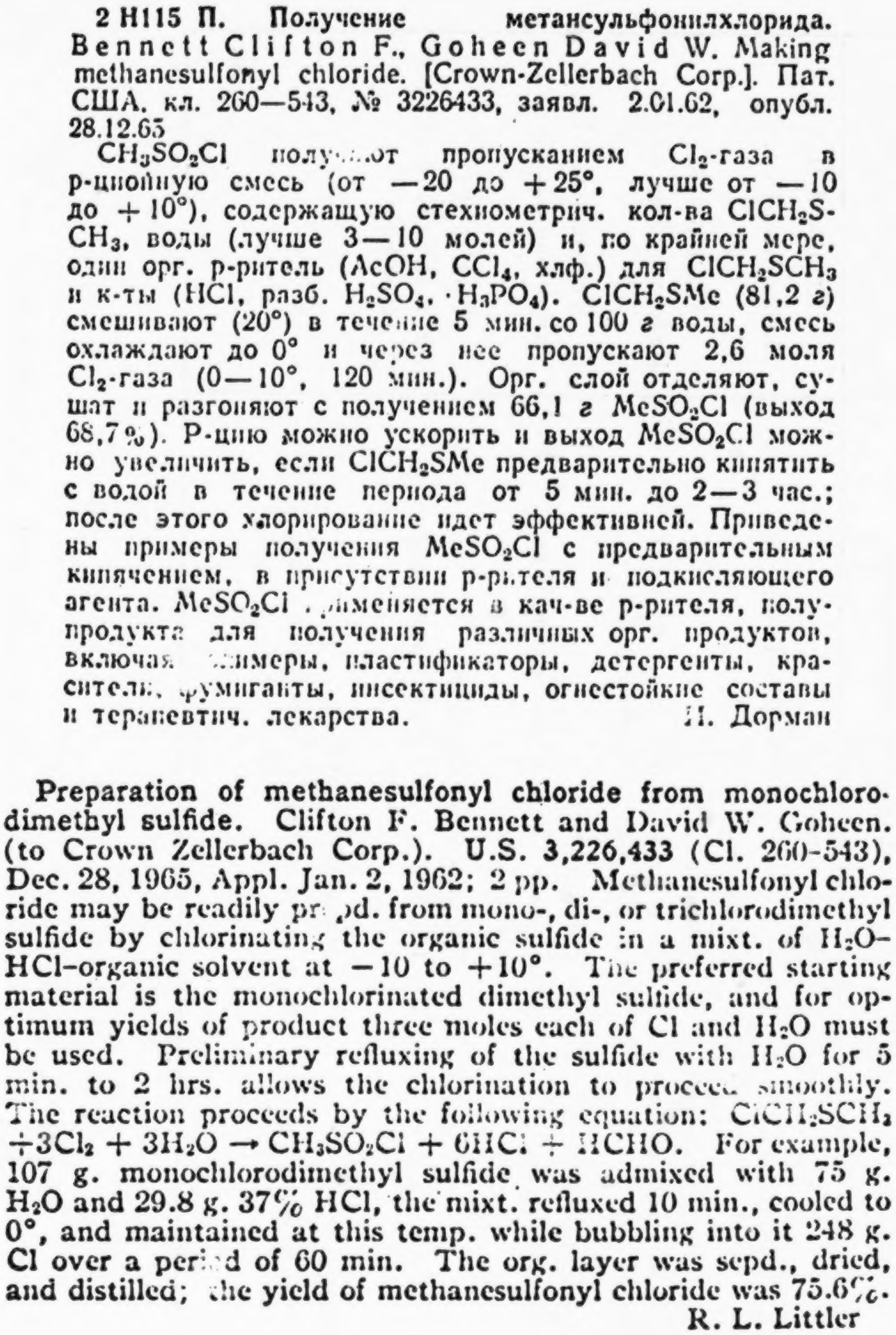


Fig. 8. Age distribution of papers cited in publications of 1965.

The horizontal axis shows the publication years of papers cited in 1965, in chronological order, the vertical axis shows the proportions of the papers of the specified publication years to the total number of references made in 1965. The dotted lines indicate the percentage of the total publications issued between 1941 and 1965 to the total number of publications cited in the publications of 1965. The diagrams are based on data taken from: Science Citation Index Guide. Philadelphia, 1966, p. 32.



2 Н115 П. Получение метансульфонилхлорида. Bennett Clifton F., Goheen David W. Making methanesulfonyl chloride. [Crown-Zellerbach Corp.]. Пат. США. кл. 260—543, № 3226433, заявл. 2.01.62, опубл. 28.12.65

CH_3SO_2Cl получают пропусканием Cl_2-газа в р-ционную смесь (от —20 до +25°, лучше от —10 до +10°), содержащую стехиометрич. кол-ва $ClCH_2SCH_3$, воды (лучше 3—10 молей) и, по крайней мере, один орг. р-ритель (AcOH, CCl_4, хлф.) для $ClCH_2SCH_3$ и к-ты (HCl, разб. H_2SO_4, $\cdot H_3PO_4$). $ClCH_2SMe$ (81,2 *г*) смешивают (20°) в течение 5 мин. со 100 *г* воды, смесь охлаждают до 0° и через нее пропускают 2,6 моля Cl_2-газа (0—10°, 120 мин.). Орг. слой отделяют, сушат и разгоняют с получением 66,1 *г* $MeSO_2Cl$ (выход 68,7%). Р-цию можно ускорить и выход $MeSO_2Cl$ можно увеличить, если $ClCH_2SMe$ предварительно кипятить с водой в течение периода от 5 мин. до 2—3 час.; после этого хлорирование идет эффективней. Приведены примеры получения $MeSO_2Cl$ с предварительным кипячением, в присутствии р-рителя и подкисляющего агента. $MeSO_2Cl$ применяется в кач-ве р-рителя, полупродукта для получения различных орг. продуктов, включая полимеры, пластификаторы, детергенты, красители, фумиганты, инсектициды, огнестойкие составы и терапевтич. лекарства. Н. Дорман

Preparation of methanesulfonyl chloride from monochlorodimethyl sulfide. Clifton F. Bennett and David W. Goheen. (to Crown Zellerbach Corp.). U.S. **3,226,433** (Cl. 260-543), Dec. 28, 1965, Appl. Jan. 2, 1962; 2 pp. Methanesulfonyl chloride may be readily prepd. from mono-, di-, or trichlorodimethyl sulfide by chlorinating the organic sulfide in a mixt. of H_2O–HCl–organic solvent at −10 to +10°. The preferred starting material is the monochlorinated dimethyl sulfide, and for optimum yields of product three moles each of Cl and H_2O must be used. Preliminary refluxing of the sulfide with H_2O for 5 min. to 2 hrs. allows the chlorination to proceed smoothly. The reaction proceeds by the following equation: $ClCH_2SCH_3 + 3Cl_2 + 3H_2O \rightarrow CH_3SO_2Cl + 6HCl + HCHO$. For example, 107 g. monochlorodimethyl sulfide was admixed with 75 g. H_2O and 29.8 g. 37% HCl, the mixt. refluxed 10 min., cooled to 0°, and maintained at this temp. while bubbling into it 248 g. Cl over a period of 60 min. The org. layer was sepd., dried, and distilled; the yield of methanesulfonyl chloride was 75.6%.
R. L. Littler

Fig. 9. Two different abstracts of the same paper as abstracted in the VINITI Abstract Journal "Chemistry" and the "Chemical Abstracts"

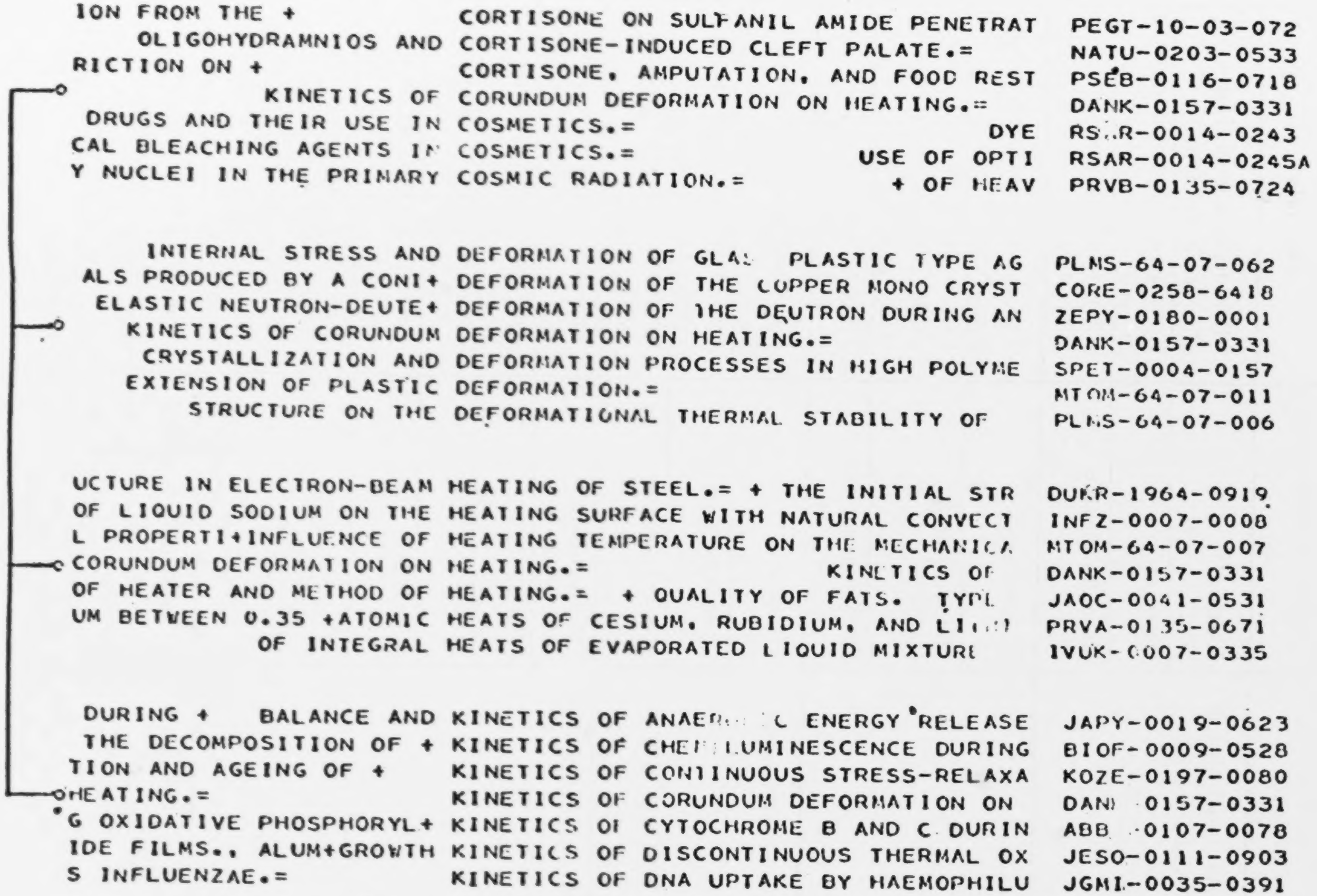
This example shows the use of the formalized chemical language for contracting the text of an original paper to its abstract. The abstracts have largely overlapping contents.

Classified order (UDC)	Alphabetical subject order	Author order
1. Phylosophy McLuhan M. Understanding media.	**Architecture** Rasmussen S.E. Towns and buildings.	**Keenan J.H.** Thermodynamics.
5. Mathematics and natural sciences Wittgenstein L. Remarks on the foundations of mathematics.	**Mathematics** Wittgenstein L. Remarks on the foundations of mathematics.	**McLuhan M.** Understanding media.
6. Applied sciences Engineering Keenan J.H. Thermodynamics.	**Paleography** Ullman B.L. Ancient writing and its influence.	**Rasmussen S.E.** Towns and buildings.
7. The arts Rasmussen S.E. Towns and buildings.	**Social communication** McLuhan M. Understanding media.	**Ullman B.L.** Ancient writing and its influence.
8. Linguistics. Philology Ullman B.L. Ancient writing and its influence.	**Thermodynamics** Keenan J.H. Thermodynamics.	

Fig. 10. Arrangement of entries for the same books in a classified, alphabetical subject and author order

Characteristic of division	Types of bibliography	Types of indexed materials
Content and type of indexed materials	General	All literature irrespective of content and type
	Special	Literature of specific content or type
	Branch-oriented	Literature in a specific subject field
Purpose	Enumerative	The widest possible coverage by content or type
	Information	Literature for specialists in a specific subject field, theme or problem
	Recommendatory	Selected literature for a specific reader category
Time of publication	Current	Currently published materials
	Retrospective	Literature published during a certain time period
	Prospective	Literature prepared for publication
Place of publication	International	Literature appearing in different languages in different countries
	National	Literature appearing in one language or country
	State	Literature published in one country
	Regional	Literature published in one locality
Level	Primary	Published works
	Secondary	Primary bibliographies
	Tertiary	Secondary bibliographies (bibliographies of bibliographies)
Method of grouping of entries	Classified	According to a scientific system of knowledge
	Subject	Alphabetically by subjects
	Author	Alphabetically by authors and titles

Fig. 11. Basic types of bibliography



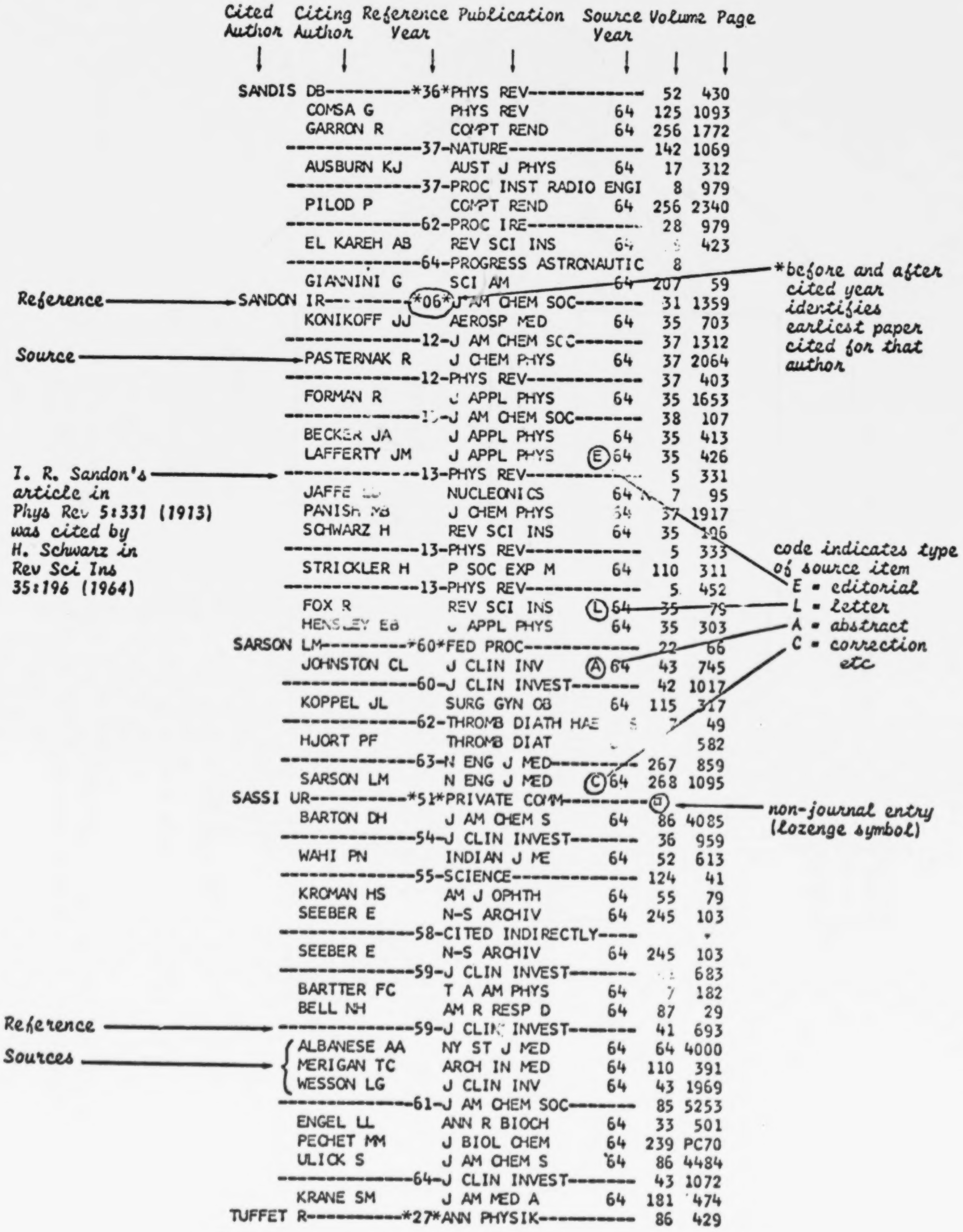
```
ION FROM THE +            CORTISONE ON SULFANIL AMIDE PENETRAT  PEGT-10-03-072
    OLIGOHYDRAMNIOS AND CORTISONE-INDUCED CLEFT PALATE.=        NATU-0203-0533
RICTION ON +              CORTISONE, AMPUTATION, AND FOOD REST  PSEB-0116-0718
          KINETICS OF CORUNDUM DEFORMATION ON HEATING.=         DANK-0157-0331
 DRUGS AND THEIR USE IN COSMETICS.=                         DYE  RSAR-0014-0243
CAL BLEACHING AGENTS IN COSMETICS.=                  USE OF OPTI RSAR-0014-0245A
Y NUCLEI IN THE PRIMARY COSMIC RADIATION.=             + OF HEAV PRVB-0135-0724


     INTERNAL STRESS AND DEFORMATION OF GLASS  PLASTIC TYPE AG  PLMS-64-07-062
 ALS PRODUCED BY A CONI+ DEFORMATION OF THE COPPER MONO CRYST   CORE-0258-6418
  ELASTIC NEUTRON-DEUTE+ DEFORMATION OF THE DEUTRON DURING AN   ZEPY-0180-0001
    KINETICS OF CORUNDUM DEFORMATION ON HEATING.=               DANK-0157-0331
     CRYSTALLIZATION AND DEFORMATION PROCESSES IN HIGH POLYME   SPET-0004-0157
     EXTENSION OF PLASTIC DEFORMATION.=                         MTOM-64-07-011
         STRUCTURE ON THE DEFORMATIONAL THERMAL STABILITY OF    PLMS-64-07-006


UCTURE IN ELECTRON-BEAM HEATING OF STEEL.= + THE INITIAL STR   DUKR-1964-0919
OF LIQUID SODIUM ON THE HEATING SURFACE WITH NATURAL CONVECT   INFZ-0007-0008
L PROPERTI+INFLUENCE OF HEATING TEMPERATURE ON THE MECHANICA   MTOM-64-07-007
CORUNDUM DEFORMATION ON HEATING.=                KINETICS OF   DANK-0157-0331
OF HEATER AND METHOD OF HEATING.=  + QUALITY OF FATS.  TYPE    JAOC-0041-0531
UM BETWEEN 0.35 +ATOMIC HEATS OF CESIUM, RUBIDIUM, AND LI[illegible]  PRVA-0135-0671
           OF INTEGRAL HEATS OF EVAPORATED LIQUID MIXTURE       IVUK-0007-0335


 DURING +   BALANCE AND KINETICS OF ANAEROBIC ENERGY RELEASE   JAPY-0019-0623
 THE DECOMPOSITION OF + KINETICS OF CHEMILUMINESCENCE DURING   BIOF-0009-0528
TION AND AGEING OF +    KINETICS OF CONTINUOUS STRESS-RELAXA   KOZE-0197-0080
HEATING.=               KINETICS OF CORUNDUM DEFORMATION ON    DANK-0157-0331
G OXIDATIVE PHOSPHORYL+ KINETICS OF CYTOCHROME B AND C DURIN   ABB -0107-0078
IDE FILMS., ALUM+GROWTH KINETICS OF DISCONTINUOUS THERMAL OX   JESO-0111-0903
S INFLUENZAE.=          KINETICS OF DNA UPTAKE BY HAEMOPHILU   JGMI-0035-0391
```

Fig. 12. Layout of entry in a permuted-title index (KWIC Index).



Cited Author
Citing Author
Reference Year
Publication
Source Year
Volume
Page
Reference
Source
I. R. Sandon's article in Phys Rev 5:331 (1913) was cited by H. Schwarz in Rev Sci Ins 35:196 (1964)
Reference
Sources
*before and after cited year identifies earliest paper cited for that author
code indicates type of source item
E = editorial
L = letter
A = abstract
C = correction
etc
non-journal entry (lozenge symbol)

```
SANDIS DB---------*36*PHYS REV-----------     52  430
       COMSA G        PHYS REV          64   125 1093
       GARRON R       COMPT REND        64   256 1772
       -------------37-NATURE-------------   142 1069
       AUSBURN KJ     AUST J PHYS       64    17  312
       -------------37-PROC INST RADIO ENGI    8  979
       PILOD P        COMPT REND        64   256 2340
       -------------62-PROC IRE-----------    28  979
       EL KAREH AB    REV SCI INS       64   [illegible]  423
       -------------64-PROGRESS ASTRONAUTIC    8
       GIANNINI G     SCI AM            64   207   59
SANDON IR---------*06*J AM CHEM SOC------     31 1359
       KONIKOFF JJ    AEROSP MED        64    35  703
       -------------12-J AM CHEM SOC------    37 1312
       PASTERNAK R    J CHEM PHYS       64    37 2064
       -------------12-PHYS REV-----------    37  403
       FORMAN R       J APPL PHYS       64    35 1653
       -------------1[illegible]-J AM CHEM SOC------    38  107
       BECKER JA      J APPL PHYS       64    35  413
       LAFFERTY JM    J APPL PHYS     (E)64   35  426
       -------------13-PHYS REV-----------     5  331
       JAFFE [illegible]      NUCLEONICS        64     7   95
       PANISH MB      J CHEM PHYS       64    37 1917
       SCHWARZ H      REV SCI INS       64    35  196
       -------------13-PHYS REV-----------     5  333
       STRICKLER H    P SOC EXP M       64   110  311
       -------------13-PHYS REV-----------     5  452
       FOX R          REV SCI INS     (L)64   35   79
       HENSLEY EB     J APPL PHYS       64    35  303
SARSON LM---------*60*FED PROC-----------     22   66
       JOHNSTON CL    J CLIN INV      (A)64   43  745
       -------------60-J CLIN INVEST------    42 1017
       KOPPEL JL      SURG GYN OB       64   115  317
       -------------62-THROMB DIATH HAE  [illegible]    7   49
       HJORT PF       THROMB DIAT                  582
       -------------63-N ENG J MED--------   267  859
       SARSON LM      N ENG J MED     (C)64  268 1095
SASSI UR----------*51*PRIVATE COMM--------◊
       BARTON DH      J AM CHEM S       64    86 4085
       -------------54-J CLIN INVEST------    36  959
       WAHI PN        INDIAN J ME       64    52  613
       -------------55-SCIENCE------------   124   41
       KROMAN HS      AM J OPHTH        64    55   79
       SEEBER E       N-S ARCHIV        64   245  103
       -------------58-CITED INDIRECTLY----
       SEEBER E       N-S ARCHIV        64   245  103
       -------------59-J CLIN INVEST------   [illegible]  683
       BARTTER FC     T A AM PHYS       64     7  182
       BELL NH        AM R RESP D       64    87   29
       -------------59-J CLIN INVEST------    41  693
      {ALBANESE AA    NY ST J MED       64    64 4000
      {MERIGAN TC     ARCH IN MED       64   110  391
      {WESSON LG      J CLIN INV        64    43 1969
       -------------61-J AM CHEM SOC------    85 5253
       ENGEL LL       ANN R BIOCH       64    33  501
       PECHET MM      J BIOL CHEM       64   239 PC70
       ULICK S        J AM CHEM S       64    86 4484
       -------------64-J CLIN INVEST------    43 1072
       KRANE SM       J AM MED A        64   181  474
TUFFET R----------*27*ANN PHYSIK----------    86  429
```

Fig. 13. Portion of a page in the Science Citation Index.

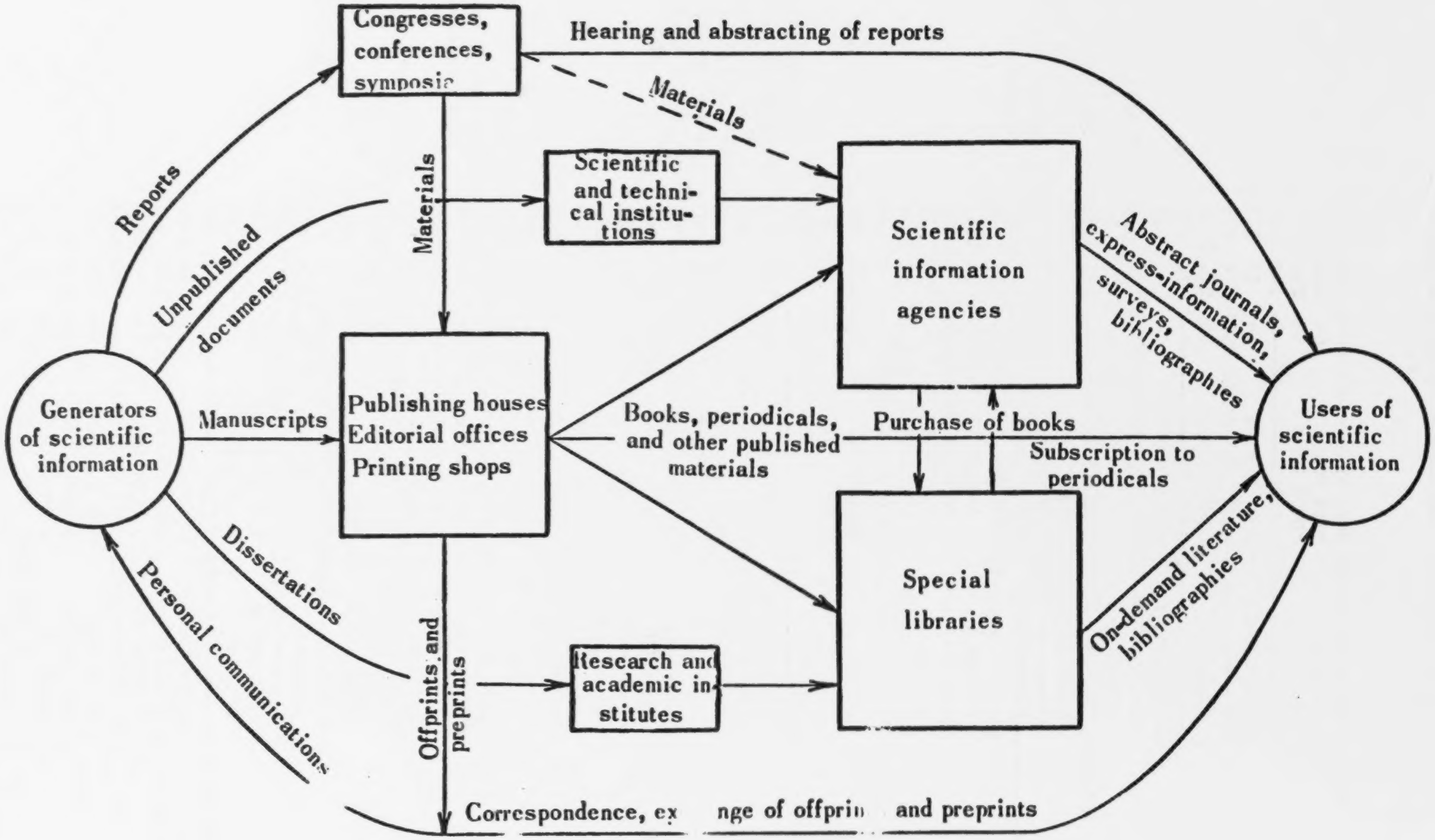

Generators of scientific information
Reports
Unpublished documents
Manuscripts
Dissertations
Personal communications
Congresses, conferences, symposia
Hearing and abstracting of reports
Materials
Materials
Scientific and technical institutions
Publishing houses Editorial offices Printing shops
Offprints and preprints
Research and academic institutes
Books, periodicals, and other published materials
Scientific information agencies
Purchase of books
Special libraries
Subscription to periodicals
Abstract journals, express-information, surveys, bibliographies
On-demand literature, bibliographies
Correspondence, ex nge of offprin and preprints
Users of scientific information

Fig. 14. Channels of dissemination of scientific information from the generators to the users

The diagram is based on the lecture "Effektivnost' deyatel'nosti organov nauchnotekhnicheskoy informatsii" (Effectiveness of activities of information agencies), by M.L. Kolchinsky (Moscow, VINITI, 1966, 29 p.).

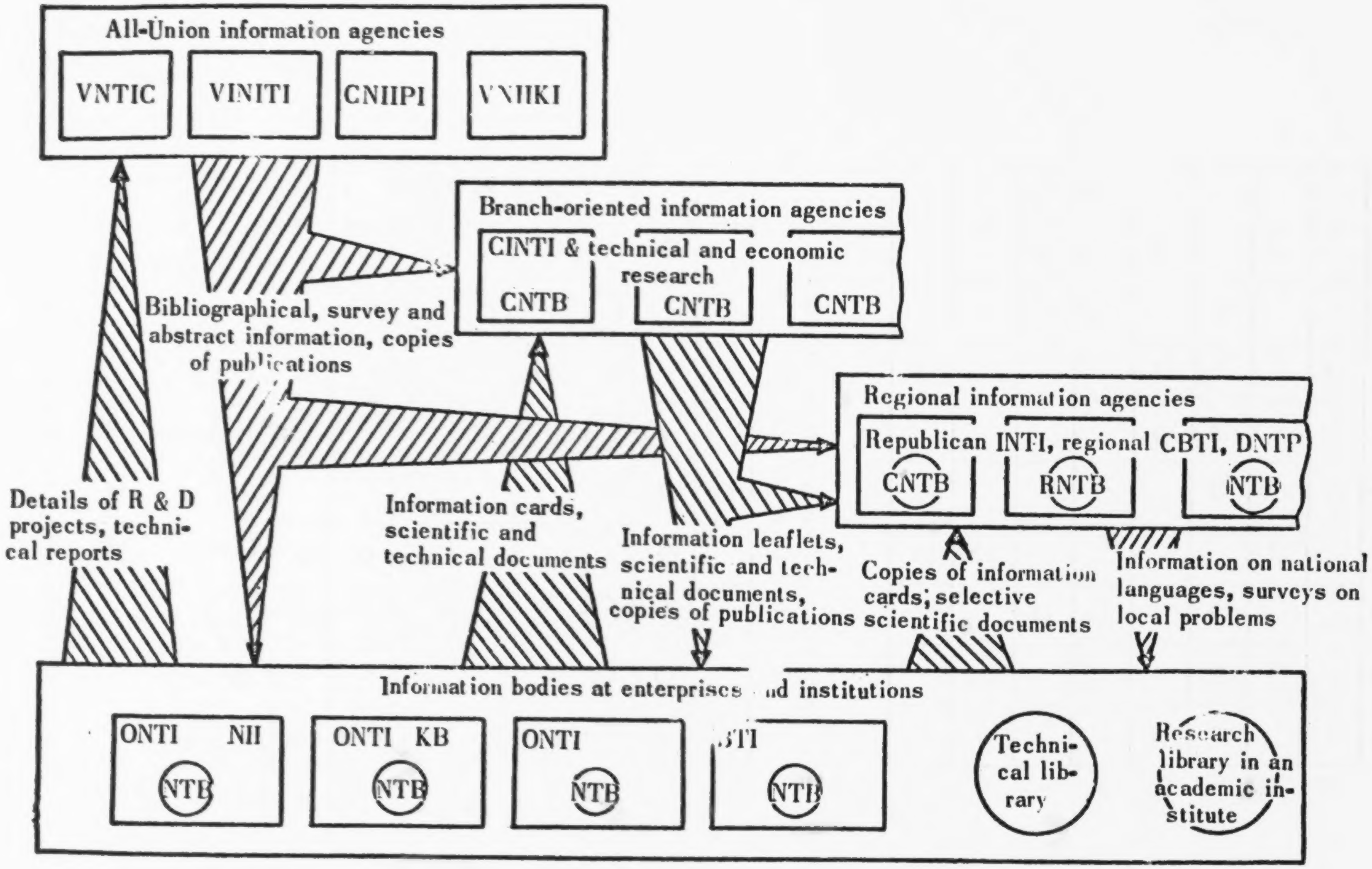
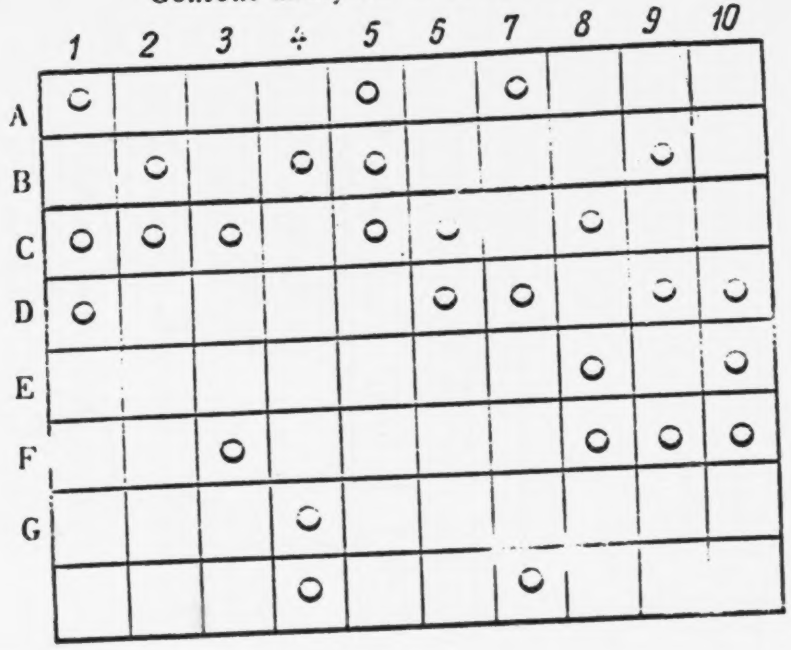

All-Union information agencies
VNTIC
VINITI
CNIIPI
VNIIKI
Branch-oriented information agencies
CINTI & technical and economic research
CNTB
CNTB
CNTB
Regional information agencies
Republican INTI, regional CBTI, DNTP
CNTB
RNTB
NTB
Details of R & D projects, technical reports
Bibliographical, survey and abstract information, copies of publications
Information cards, scientific and technical documents
Information leaflets, scientific and technical documents, copies of publications
Copies of information cards; selective scientific documents
Information on national languages, surveys on local problems
Information bodies at enterprises and institutions
ONTI NII
NTB
ONTI KB
NTB
ONTI
NTB
BTI
NTB
Technical library
Research library in an academic institute

Fig. 15. Distribution of scientific and technical information flows among the information agencies in the USSR

Content analysis of documents


1
2
3
4
5
6
7
8
9
10
A
B
C
D
E
F
G

Searching for ducuments by characteristics
[C and D] or [C and F]

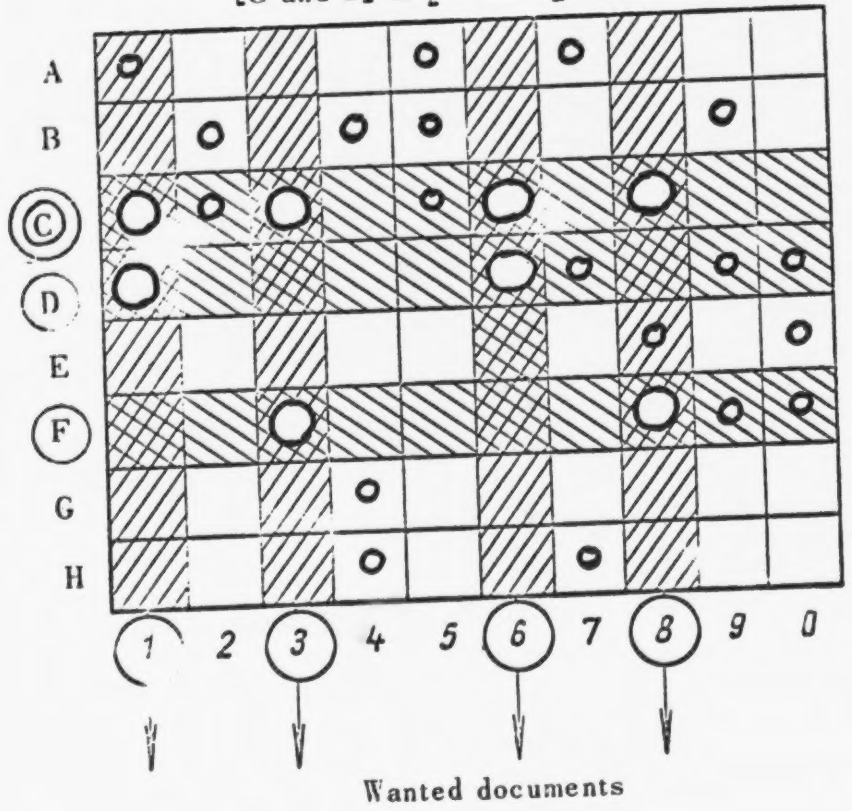

A
B
C
D
E
F
G
H
1
2
3
4
5
6
7
8
9
0
Wanted documents

Fig. 16. Diagram of information retrieval.

Used on the paper by B.-A. Lipetz "Information storage and retrieval" in "Scientific American", 1966, vol. 215, № 3, p. 226.

Documents	Relevant	Non-relevant	
Retrieved	a	b	$a+b$
Non-retrieved	c	d	$c+d$
	$a+c$	$b+d$	

Recall $\frac{a}{a+c}\cdot 100\%$

Precision $\frac{a}{a+b}\cdot 100\%$

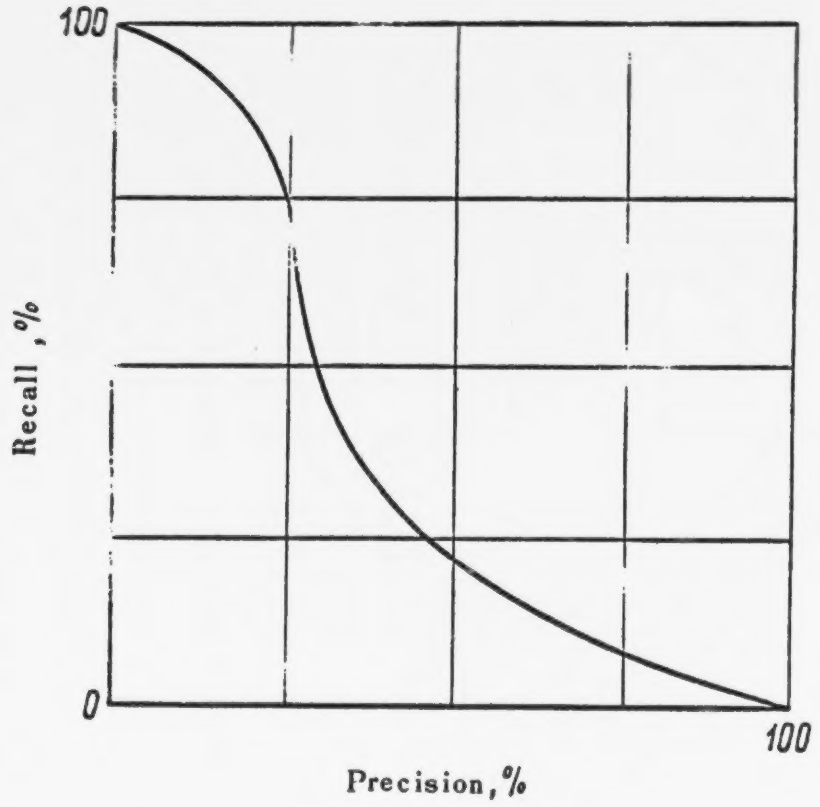


Fig. 17. Dependence of recall on precision (after C. Cleverdon)

C.W. Cleverdon and J. Mills. The testing of index language divices. – "Aslib Proceedings", 1963, v. 15, № 4, p. 106–130.

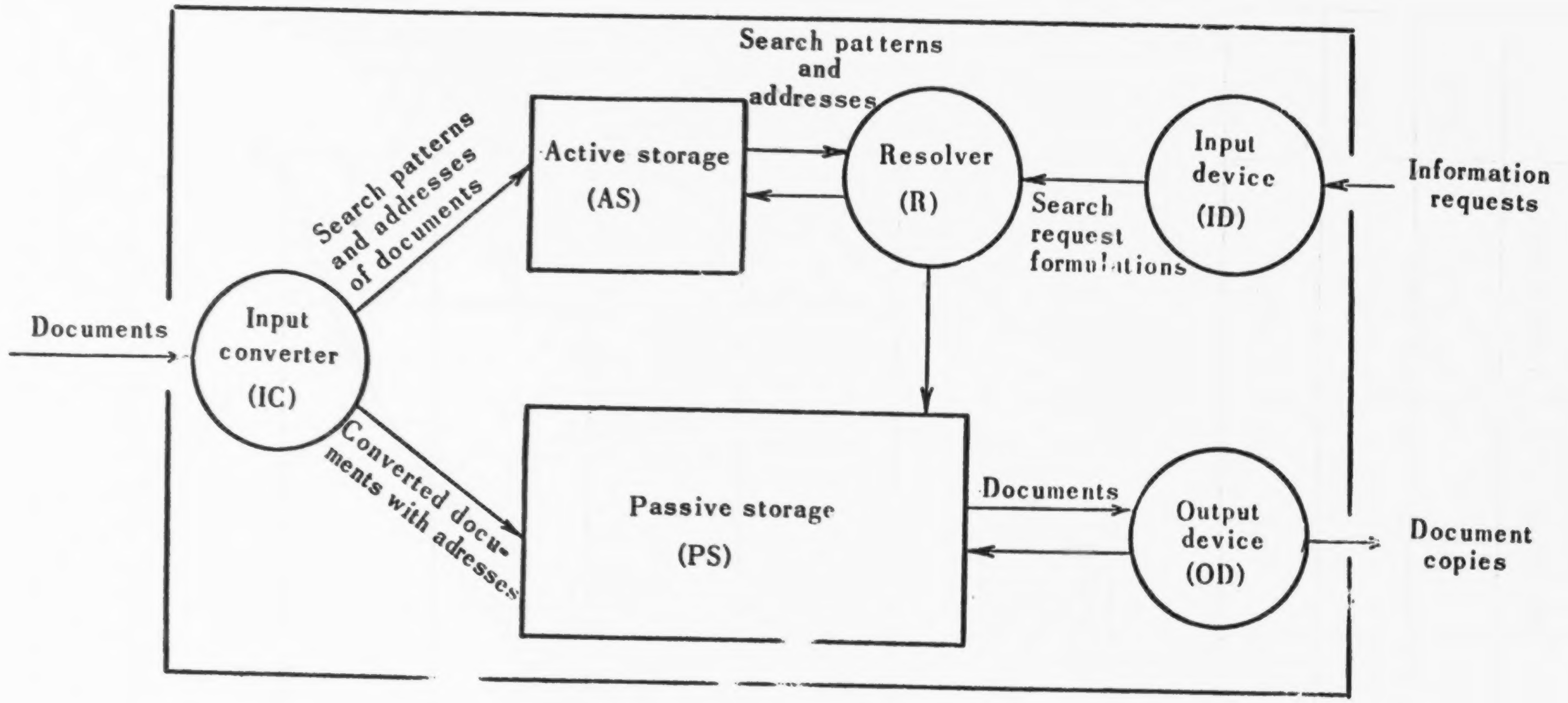

Documents
Input converter (IC)
Search patterns and addresses of documents
Converted documents with adresses
Active storage (AS)
Passive storage (PS)
Search patterns and addresses
Resolver (R)
Search request formulations
Input device (ID)
Information requests
Documents
Output device (OD)
Document copies

Fig. 18. General block-diagram of an information retrieval system.

Based on the paper, by G.E. Vleduts, "O nekotorykh storonakh issledovaniy po sozdaniyu informatsionno-poiskovykh sistem" (Some aspects of research in the development of information retrieval systems) in "Nauchno-Tekhnicheskaya Informatsiya", 1961, № 1, p. 32.

Specimen title-page	Bibliographical particulars
International Series of Monographs in Library and Information Science Volume 10	Information before the title
AUTOMATION IN LIBRARIES	Title
by **Richard T. Kimber** Lecturer, School of Library and Information Studies, The Queen's University of Belfast	Subtitle information
PERGAMON PRESS Oxford. New York. Toronto Sydney. Braunschweig	Imprint

Specimen entry

Kimber R.T.
Automation in libraries. By Richard T. Kimber
Oxford /a.o./, Pergamon, /1970/.
VIII, 140 p., fig. (Int. ser. of monographs in libr. and information science. Vol. 10).
Bibliogr. p. 134–138.

Fig. 19. Arrangement of bibliographical particulars on the title page and in the entry

Publication types			Entries	
			Main	Added
Authorship		1. Works of one, two or three authors	Surname (forename) of first author	Added entries under second and third authors
		2. Works of four or more authors	Title	Added entry under first author
		3. Collections of works of different authors		Added entry under compiler (editor)
		4. Official publications and publications with standard title	Name of institution (copporate author)	Added entry under title
Number of volumes and frequency of appearance	Books	5. Multivolume publications	Series entry for the whole publication	Entries for individual volumes
		6. [illegible] publications	Entries for each issue in a series	Series entry for the whole publication
	Periodicals	7. Continued publications (proceedings, etc.)	Series entry for the whole publication under the general title	Entries for separate volumes under the author and particular title
		8. Journals, magazines		Entries for individual issues under the general title
		9. Newspapers		
Polygraphical autonomy		10. Parts of publications (articles from collections and periodicals)	Analytical entry under author or title of the article	–
		11. Parts (chapters) of publications	Analytical entry under author or title of the whole work	–
		12. Reviews	Entry under author or title of the reviewed work	Entry under the author or title of the review

Fig. 20. Scheme of rules for description of the basic types of publications

1	Tripathi S.M. Modern cataloguing. Theory and practice. Agra, Agarwala, /1969/. XII, 289 p. Bibliogr. p. 287–289.
2	A Handbook of the mosquitoes of the Southeastern United States. Washington, U.S. Gov. print. off., 1960. 188 p., ill. Aut.: W.V. King, G.H. Brandley, C.N. Smith and W.C. McDuffie.
3	Computer based information retrieval systems. Ed. by B. Houghton. London, Bingley, 1968. 136 p., ill.
4	Unesco. General catalogue of Unesco publications and Unesco sponsored publications. 1946–1959. Paris, UNESCO, 1962. XVI, 217 p.
5	International forum on informatics. Volume of papers. Vol. 1–2. Moscow, 1969. (All-Union Inst. for Scient. and Techn. Information). Vol. 1. 656 p. Vol. 2. 603 p.
6	Kimber R.T. Automation in libraries. Oxford /a.o./, Pergamon, 1970. VIII, 140 p., fig. (Intern. ser. of monographs in libr. and information science. Vol. 10).
7	Royal Society of Edinburgh. Transactions... Edinburgh–London, 1954–. Vol. 62. № 1–2. 1954. Vol. 63. № 1–3. 1955–1959. Vol. 64. № 1–18. 1959–1961.
8	Unesco Bulletin for Libraries. Vol. 23–24. Paris, 1969–1970. 1969. Vol. 23, № 1–6. 1970. Vol. 24, № 1–3.
9	Financial [illegible] industry, commerce, public affairs. London, 1958–. 1958 1959 1960
10	Price D.J. de S. Networks of scientific papers. – "Science", 1965, vol. 149, № 3683, p. 514.
11	Coates E.J. Subject catalogues. Headings and structure. London, The Libr. Assoc., 1960. Chap. 13. Use and search strategy. P. 159–173.
12	On theoretical problems of informatics. Moscow, All-Union Inst. for Scient. and Techn. Information, 1969. Review: Brookes B.C. – "Journal of Documentation", 1969, vol. 25, № 3, p. 262–264.

Fig. 21. Specimen index entries.
(Serial numbers correspond to the chart in Fig. 20)

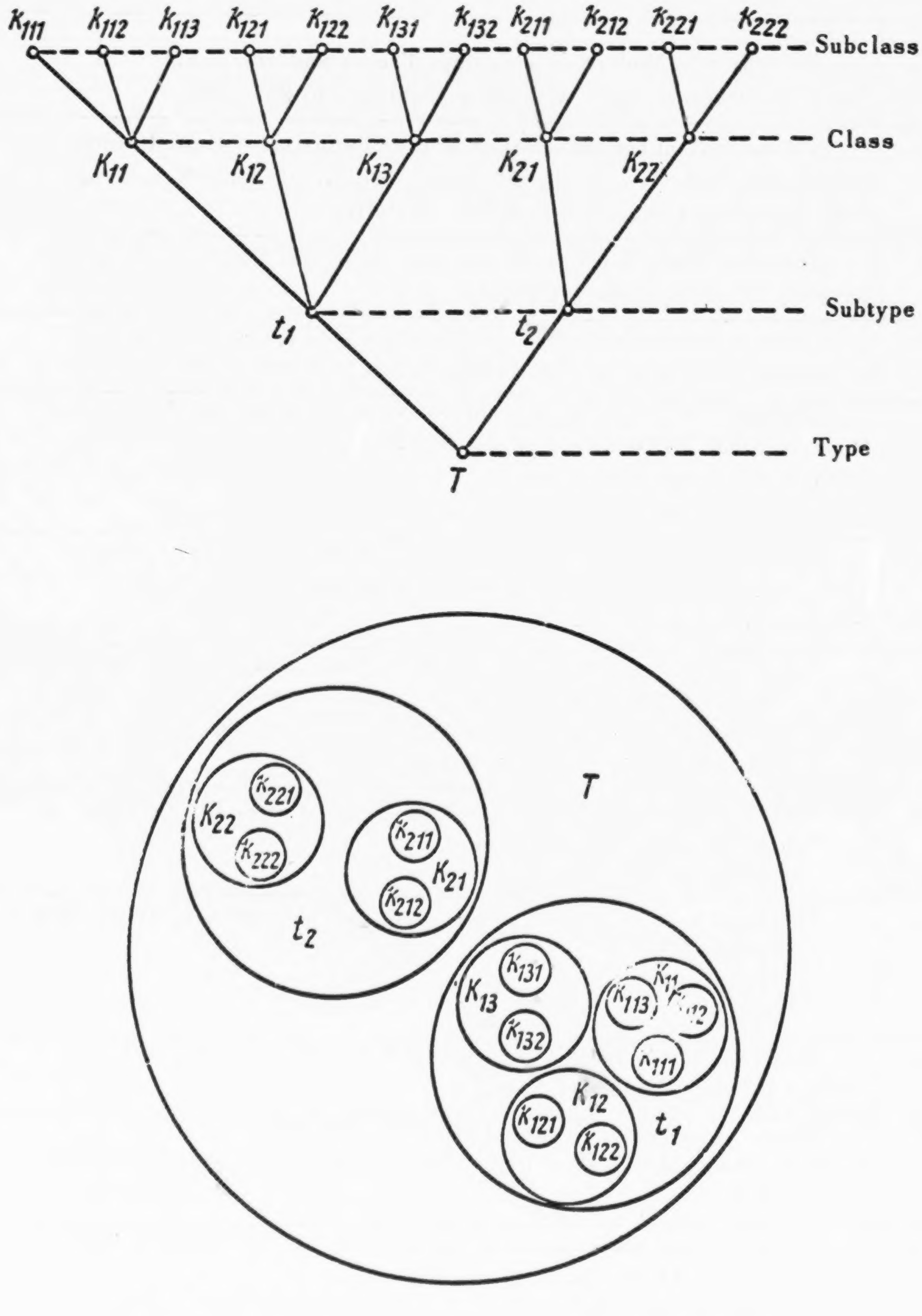

K_{111}
K_{112}
K_{113}
K_{121}
K_{122}
K_{131}
K_{132}
K_{211}
K_{212}
K_{221}
K_{222}
Subclass
K_{11}
K_{12}
K_{13}
K_{21}
K_{22}
Class
t_1
t_2
Subtype
T
Type
T
K_{22}
K_{221}
K_{222}
K_{211}
K_{212}
K_{21}
t_2
K_{131}
K_{13}
K_{132}
K_{11}
K_{113}
K_{112}
K_{111}
K_{12}
K_{121}
K_{122}
t_1

Fig. 22. Graph of an hierarchical classification and the relationship between its divisions (represented by means of Euler diagrams)

0	**GENERALITIES OF KNOWLEDGE**
00	Prolegomena. Fundamentals of knowledge and culture
01	Bibliography. Catalogues and booklists
02	Librarianship and reading
03	Encyclopaedias. Dictionaries. Reference books
04	Pamphlets, cuttings, contributions, summary publications
05	Periodicals. Reviews. Serial publications
06	Organizations. Associations. Conventions. Museums
07	Newspapers. The press. Journalism
08	Collective publications and documents of special form
09	Manuscripts. Rare, remarkable books and publications
1	**PHILOSOPHY. METAPHYSICS. PSYCHOLOGY. LOGIC. ETHICS**
11	Metaphysics and ontology
12	Special metaphysics
13	Spiritual life. Cultural and moral systems, spheres
14	Philosophical systems, speculations and theories
15	[Unoccupied, except for 159.9]
16	Logic and reasoning. Theory of knowledge
17	Ethics. Morals. Convention
18	Aesthetics and taste (in general)
19	History of philosophy
2	**RELIGION. THEOLOGY**
21	Natural theology
22	Holy Scripture. The Bible
23	Dogmatic theology. Apologetics
24	Practical theology. The religious life. Christian living
25	Pastoral theology
26	The Christian Church in general
27	Christian Church history
28	Christian Churches, worshipping bodies and sects
29	Oriental and other non-Christian religions and cults
3	**SOCIAL SCIENCES**
30	Sociology. Sociography
31	Statistics. Demography
32	Political science. Politics. Current affairs
33	Political economy. Economics
34	Jurisprudence. Law. Legislation
35	Public administration. Government. Defence
36	Social relief, welfare, movements. Insurance
37	Education. Teaching. Training. Schools
38	Commerce. Trade. Communications and transport
39	Ethnography. Customs and rites. Folklore
4	[Unoccupied, since 1962]

Fig. 23. Main divisions in the Universal Decimal Classification (U.D.C.).

5	**MATHEMATICS AND NATURAL SCIENCES**
51	Mathematics
52	Astronomy. Geodesy and surveying
53	Physics and mechanics
54	Chemistry
55	Geology and associated sciences. Meteorology
56	Palaeontology. Fossils
57	Biological sciences
58	Botany. Plant biology and taxonomy
59	Zoology. Animal biology and taxonomy
6	**APPLIED SCIENCES. MEDICINE. TECHNOLOGY**
61	Medical sciences. Health and safety
62	Engineering and technology generally
63	Agriculture. Forestry. Livestock. Fisheries
64	Domestic science and economy. Household management
65	Management. Organization of industry, business, communication and transport
66	Chemical industry and technology
67	Industries and crafts based on processable materials
68	Specialized trades and industries for finished articles
69	Building: materials, construction, trades
7	**THE ARTS. RECREATION. ENTERTAINMENT. SPORT**
71	Physical planning. Landscape, parks and gardens
72	Architecture
73	Sculpture and the plastic arts
74	Drawing and minor (decorative) arts or crafts
75	Painting
76	Engraving, prints and (commercial) graphic art
77	Photography and cinematography
78	Music
79	Entertainment. Pastimes. Games. Sport
8	**LINGUISTICS. PHILOLOGY. BELLES-LETTRES. LITERATURE**
80	General linguistics and philology
82/89	Literatures of individual languages
9	**GEOGRAPHY. BIOGRAPHY. HISTORY**
91	Geography. Exploration. Travel
92	Biography
93	History
94/99	Mediaeval and modern history
	SYNOPSIS OF AUXILIARIES (IN ORDER OF FILING)
a.	Aggregation signs + and / (*preceding the simple class number*)
b.	Relation signs : and [ ]
c.	Language auxiliaries = ...
d.	Form and presentation auxiliaries (0...)
e.	Place, region, country auxiliaries (1/9)
f.	Race, people, nationality auxiliaries (=...)
g.	Time, date, period auxiliaries "..." (*alternatively before* e.)
h.	Alphabetical and numerical (non-UDC) auxiliaries A/Z, No. 1 to ...
i.	Point of view subdivisions .00...
k.	Special auxiliaries – ... and .0...

Fig. 24. Main divisions (continued), and synopsis of auxiliaries in the U.D.C.

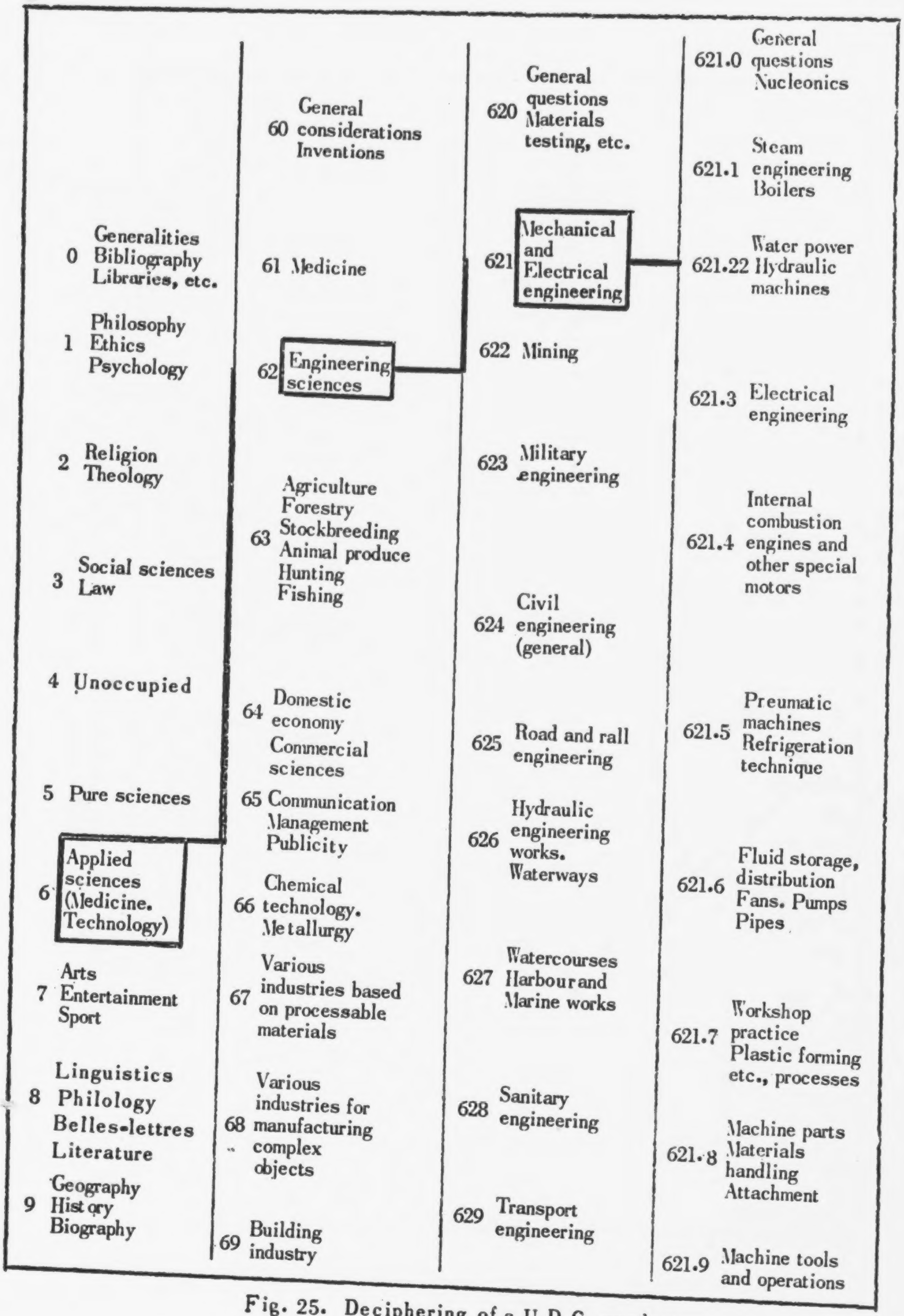

0 Generalities Bibliography Libraries, etc.
1 Philosophy Ethics Psychology
2 Religion Theology
3 Social sciences Law
4 Unoccupied
5 Pure sciences
6 Applied sciences (Medicine. Technology)
7 Arts Entertainment Sport
8 Linguistics Philology Belles-lettres Literature
9 Geography History Biography
60 General considerations Inventions
61 Medicine
62 Engineering sciences
63 Agriculture Forestry Stockbreeding Animal produce Hunting Fishing
64 Domestic economy Commercial sciences
65 Communication Management Publicity
66 Chemical technology. Metallurgy
67 Various industries based on processable materials
68 Various industries for manufacturing complex objects
69 Building industry
620 General questions Materials testing, etc.
621 Mechanical and Electrical engineering
622 Mining
623 Military engineering
624 Civil engineering (general)
625 Road and rall engineering
626 Hydraulic engineering works. Waterways
627 Watercourses Harbour and Marine works
628 Sanitary engineering
629 Transport engineering
621.0 General questions Nucleonics
621.1 Steam engineering Boilers
621.22 Water power Hydraulic machines
621.3 Electrical engineering
621.4 Internal combustion engines and other special motors
621.5 Preumatic machines Refrigeration technique
621.6 Fluid storage, distribution Fans. Pumps Pipes
621.7 Workshop practice Plastic forming etc., processes
621.8 Machine parts Materials handling Attachment
621.9 Machine tools and operations

Fig. 25. Deciphering of a U.D.C. number

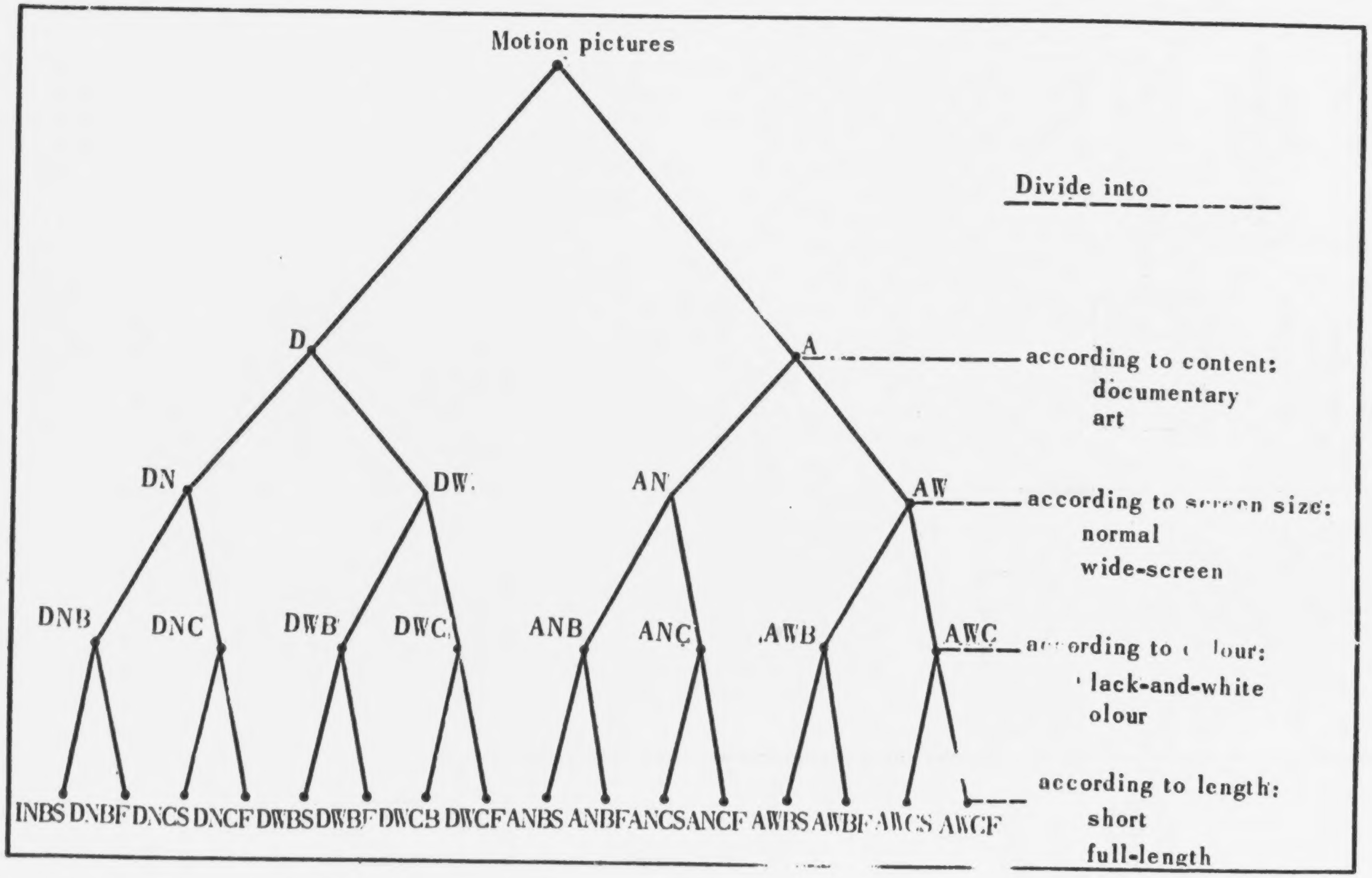

Motion pictures
Divide into
D
A
according to content:
documentary
art
DN
DW
AN
AW
according to screen size:
normal
wide-screen
DNB
DNC
DWB
DWC
ANB
ANC
AWB
AWC
according to colour:
black-and-white
colour
according to length:
short
full-length
DNBS
DNBF
DNCS
DNCF
DWBS
DWBF
DWCS
DWCF
ANBS
ANBF
ANCS
ANCF
AWBS
AWBF
AWCS
AWCF

Fig. 26. A hypothetical classification of objects according to attributes wich do not form natural hierarchy.

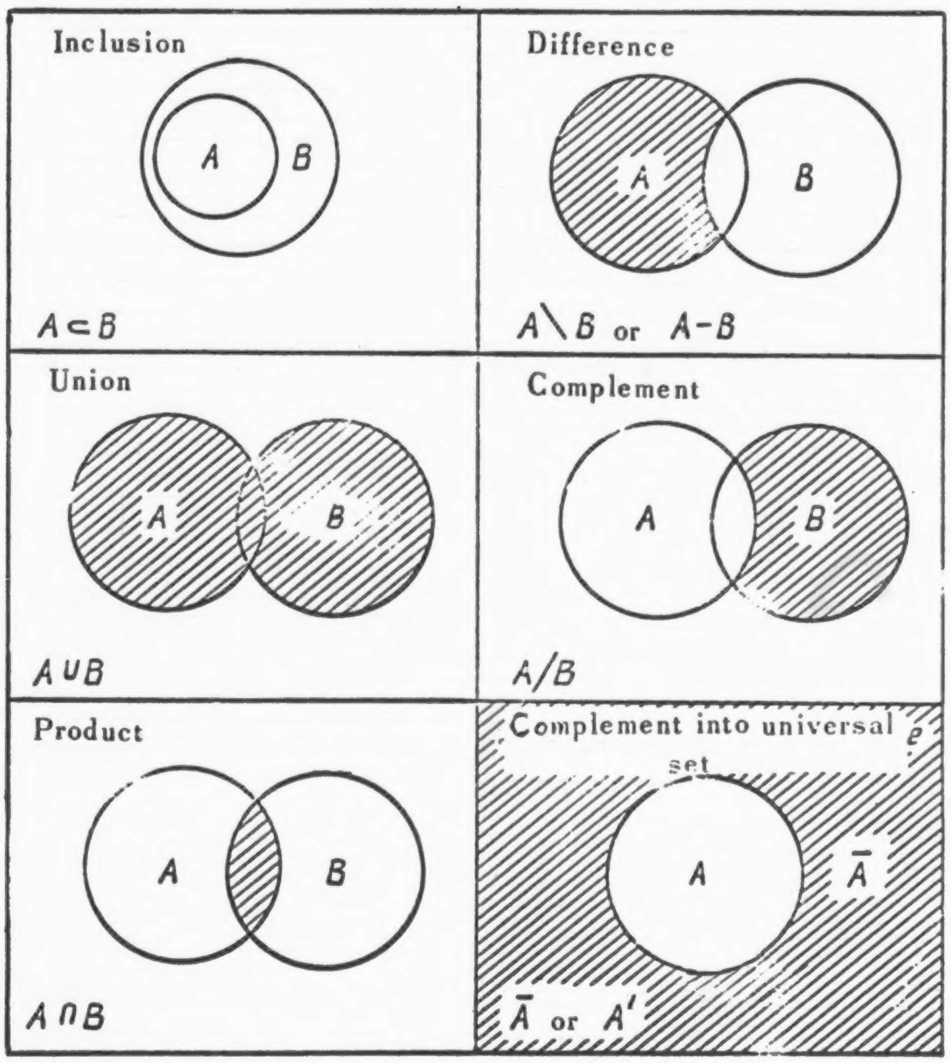

Inclusion
A
B
$A \subset B$
Difference
A
B
$A \setminus B$ or $A-B$
Union
A
B
$A \cup B$
Complement
A
B
A/B
Product
A
B
$A \cap B$
Complement into universal set
e
A
$\bar{A}$
$\bar{A}$ or A'

Fig. 27. Graphical representation of relations between sets (by means of Venn diagrams)

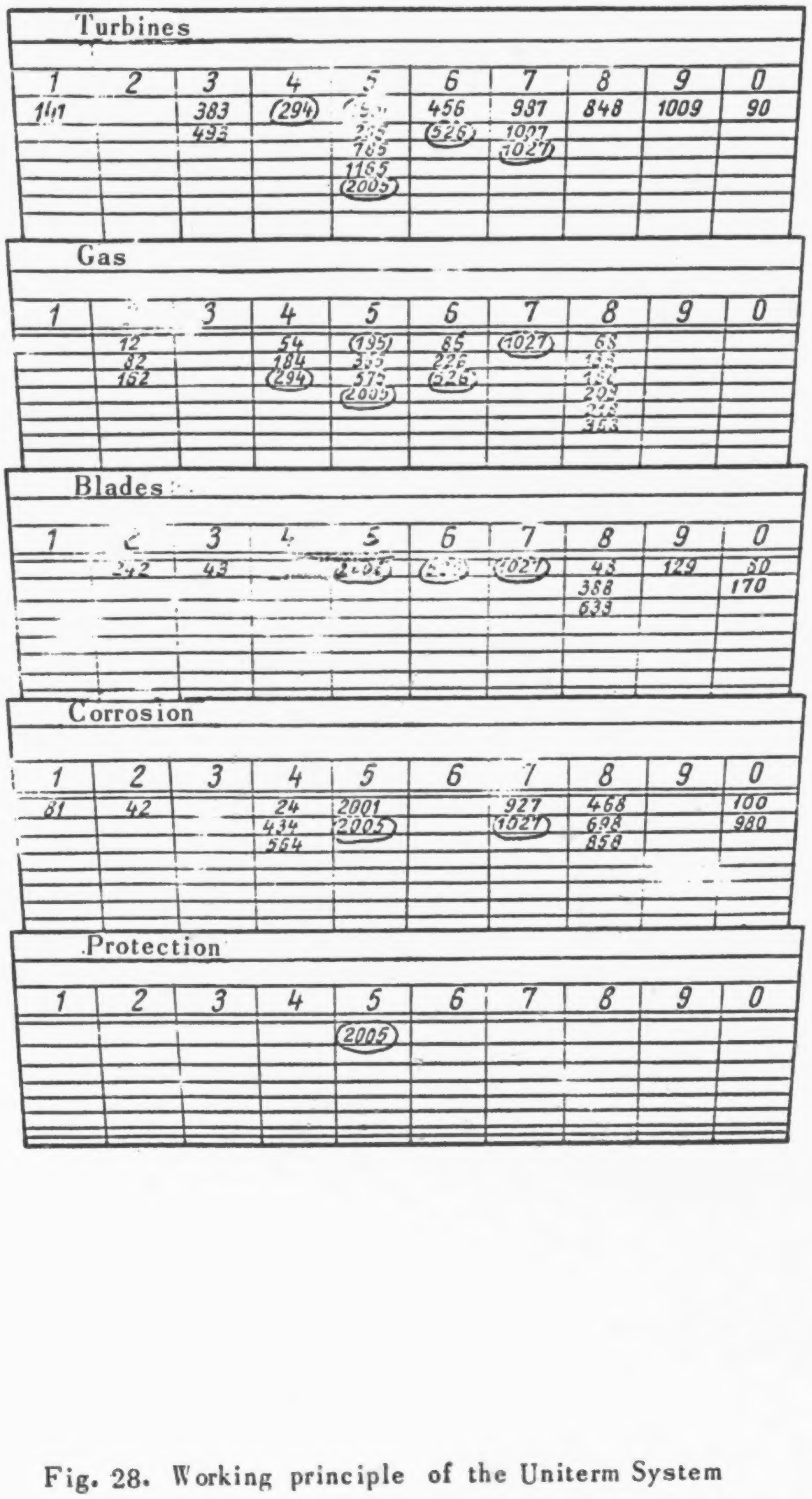

Turbines
1 2 3 4 5 6 7 8 9 0
141
383
493
294
785
1165
2005
456
526
987
1007
1027
848
1009
90
Gas
1 3 4 5 6 7 8 9 0
12
82
152
54
184
294
195
575
2005
85
226
526
1027
68
Blades
1 2 3 4 5 6 7 8 9 0
242
43
1027
43
388
633
129
170
Corrosion
1 2 3 4 5 6 7 8 9 0
81
42
24
434
564
2001
2005
927
1027
468
698
858
100
980
Protection
1 2 3 4 5 6 7 8 9 0
2005

Fig. 28. Working principle of the Uniterm System

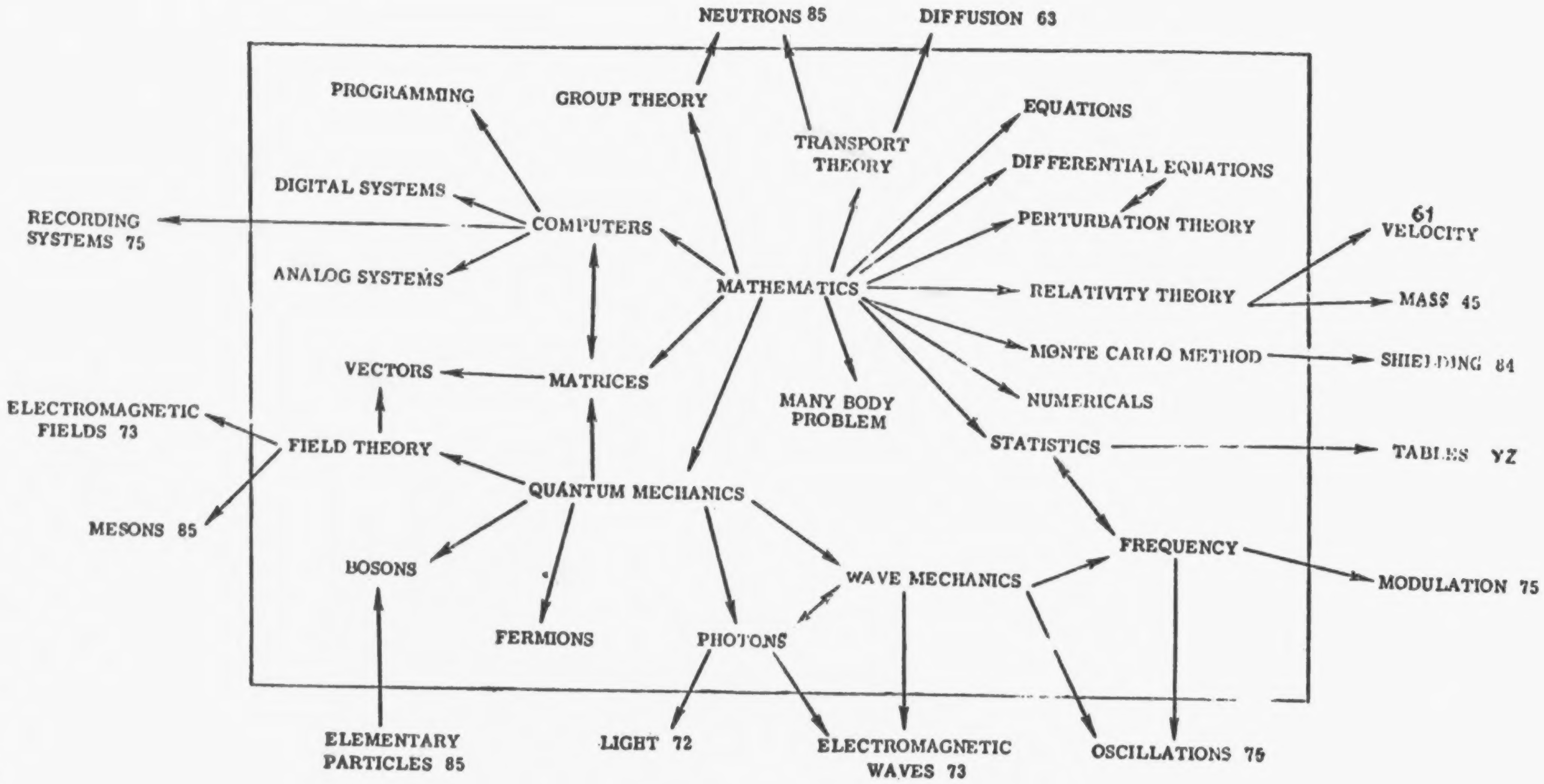

NEUTRONS 85
DIFFUSION 63
PROGRAMMING
GROUP THEORY
TRANSPORT THEORY
EQUATIONS
DIGITAL SYSTEMS
DIFFERENTIAL EQUATIONS
RECORDING SYSTEMS 75
COMPUTERS
PERTURBATION THEORY
61 VELOCITY
ANALOG SYSTEMS
MATHEMATICS
RELATIVITY THEORY
MASS 45
VECTORS
MATRICES
MONTE CARLO METHOD
SHIELDING 84
ELECTROMAGNETIC FIELDS 73
MANY BODY PROBLEM
NUMERICALS
FIELD THEORY
STATISTICS
TABLES YZ
QUANTUM MECHANICS
MESONS 85
FREQUENCY
BOSONS
WAVE MECHANICS
MODULATION 75
FERMIONS
PHOTONS
ELEMENTARY PARTICLES 85
LIGHT 72
ELECTROMAGNETIC WAVES 73
OSCILLATIONS 75

Fig. 29. Specimen layout of a subject class in a thesaurus

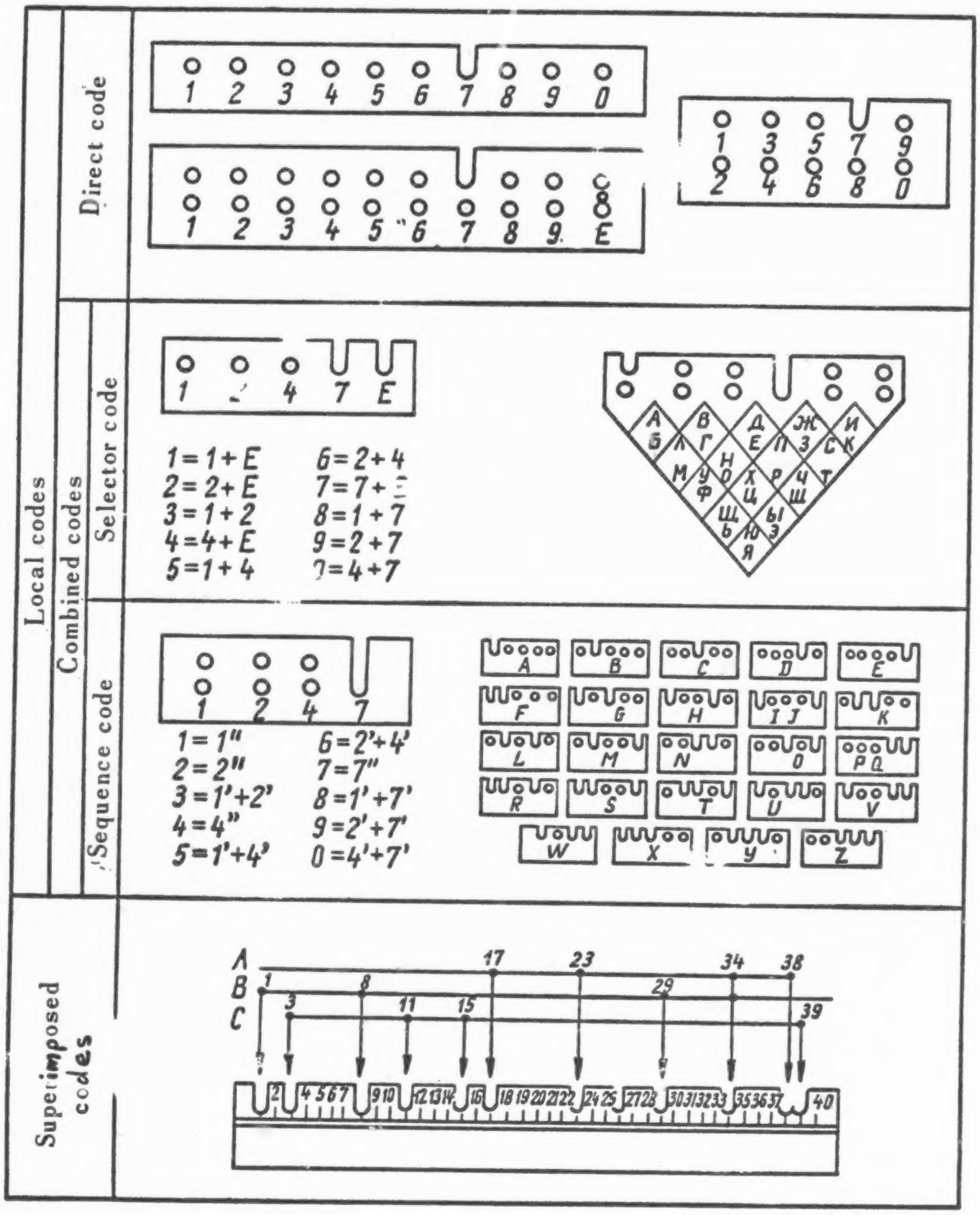

Local codes
Direct code
Combined codes
Selector code
Sequence code
Superimposed codes
1 2 3 4 5 6 7 8 9 0
1 2 3 4 5 6 7 8 9 E
1 3 5 7 9
2 4 6 8 0
1 2 4 7 E
1=1+E 6=2+4
2=2+E 7=7+E
3=1+2 8=1+7
4=4+E 9=2+7
5=1+4 0=4+7
А Б В Г Д Е Ж З И К Л М Н О П Р С Т У Ф Х Ц Ч Ш Щ Ы Ь Э Ю Я
1 2 4 7
1=1" 6=2'+4'
2=2" 7=7"
3=1'+2' 8=1'+7'
4=4" 9=2'+7'
5=1'+4' 0=4'+7'
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
A B C
1 3 8 11 15 17 23 29 34 38 39
2 4 5 6 7 9 10 12 13 14 16 18 19 20 21 22 24 25 27 28 30 31 32 33 35 36 37 40

Fig. 30. Examples of codes for edge-punched cards.

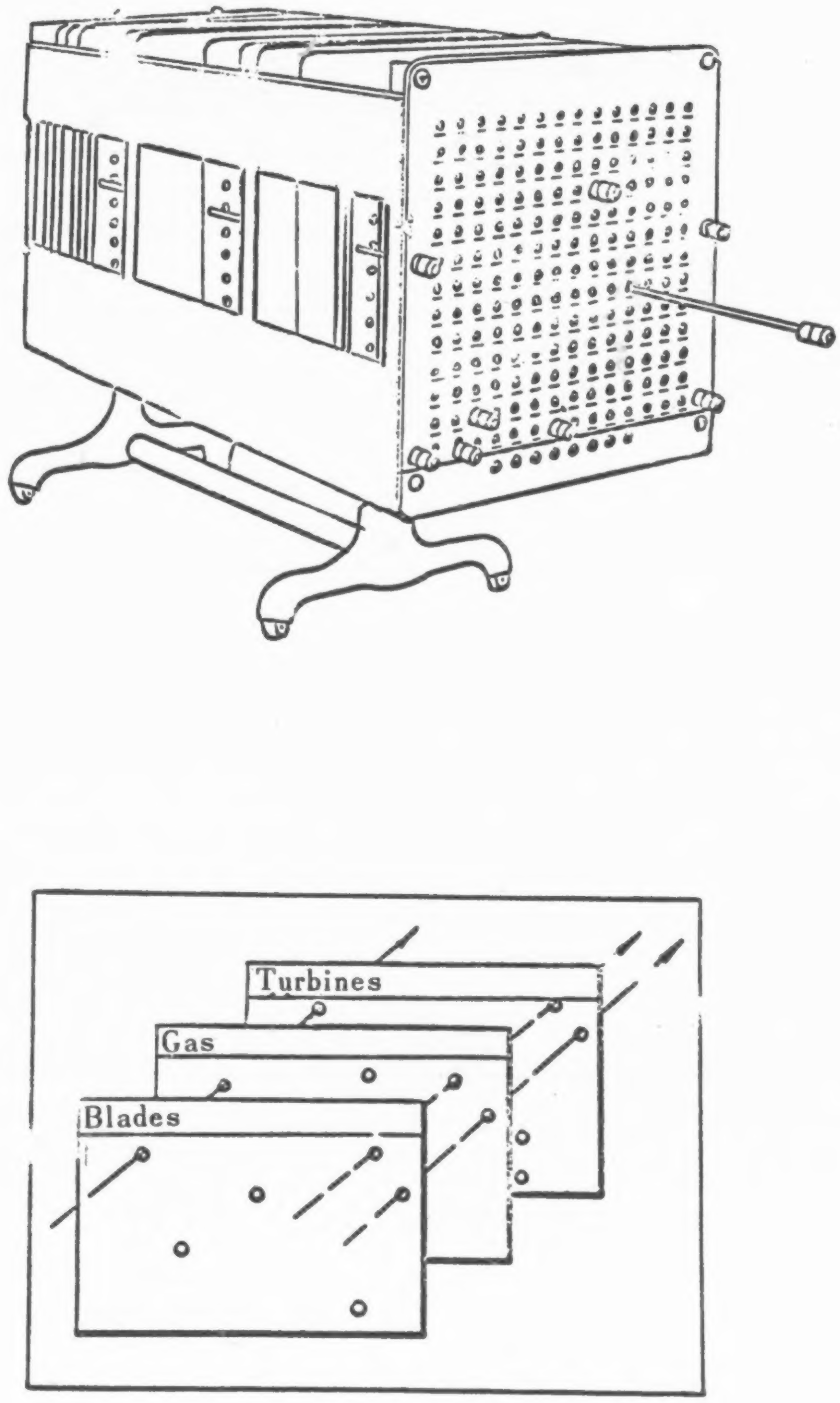

Turbines
Gas
Blades

Fig. 31. The working principle of body-punched (top) and superimposable (botton) cards.

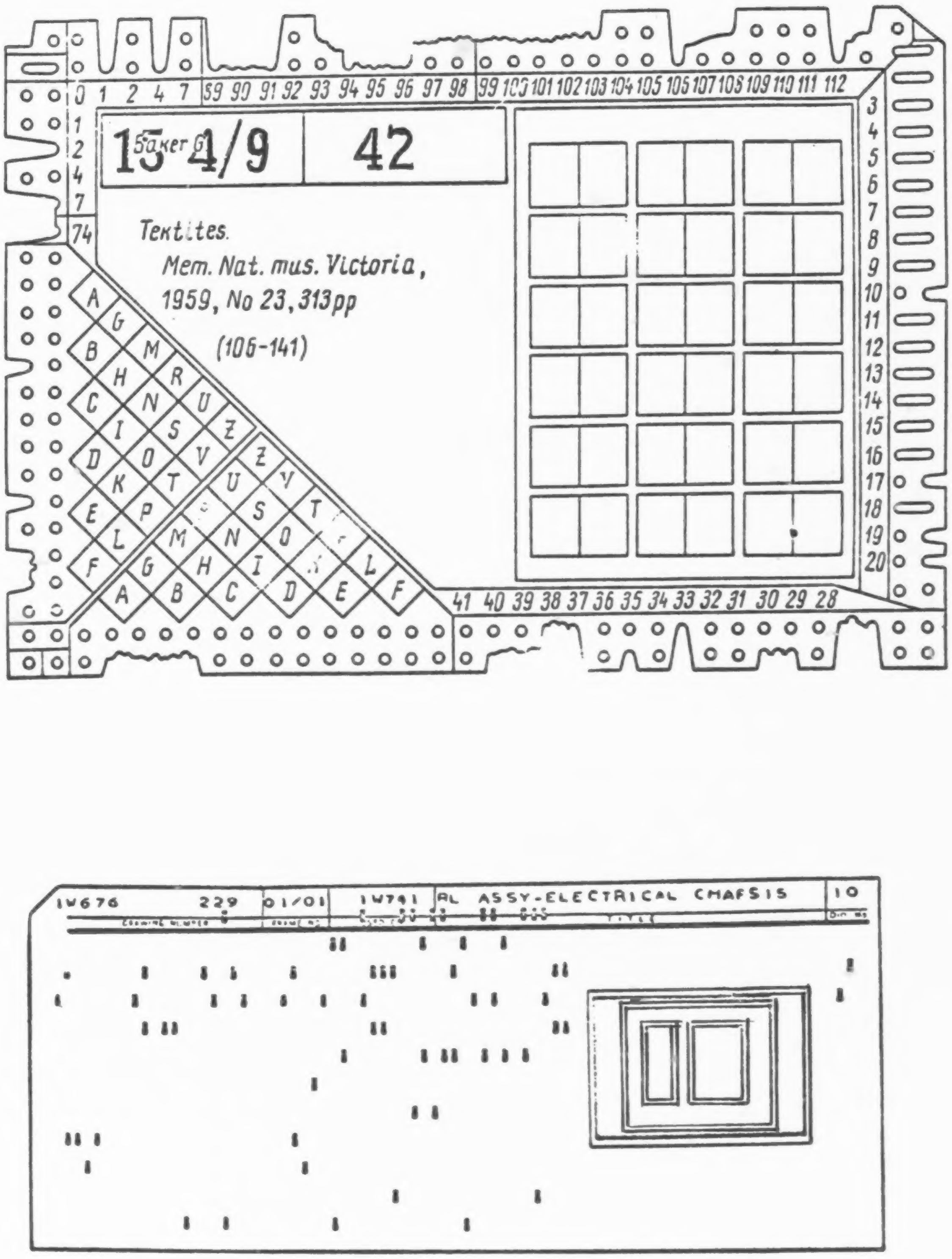

0 1 2 4 7 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112
Baker G
15 4/9
42
Tektites.
Mem. Nat. mus. Victoria,
1959, No 23, 313pp
(106-141)
A B C D E F G H I K L M N O P R S T U V Z
41 40 39 38 37 36 35 34 33 32 31 30 29 28
1W676 229
01/01
1W791
AL ASSY-ELECTRICAL CHAFSIS
10

Fig. 32. Aperture cards (edge-punched card and mechanically sorted card)

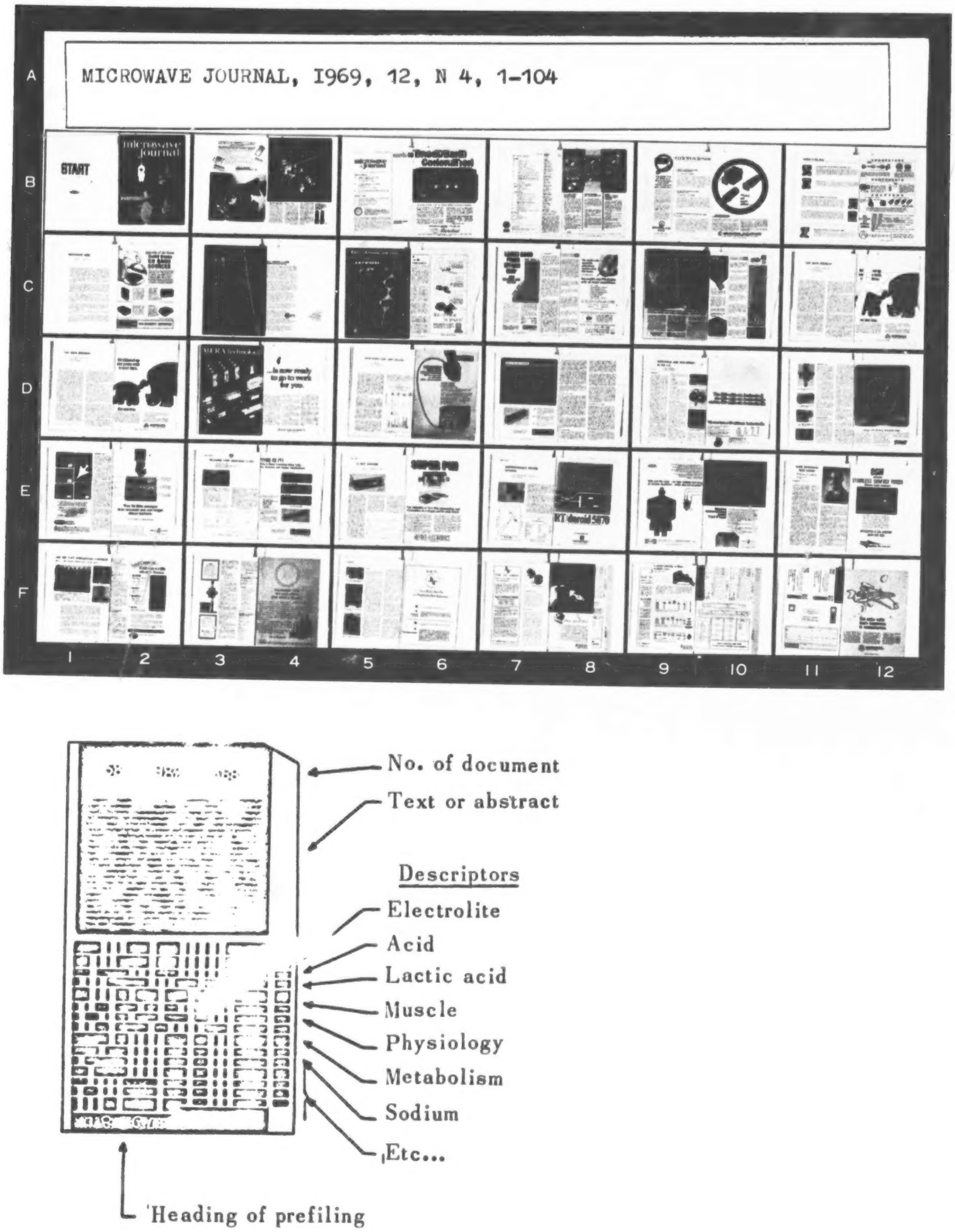

A
MICROWAVE JOURNAL, I969, 12, N 4, 1-104
B
C
D
E
F
START
1
2
3
4
5
6
7
8
9
10
11
12
No. of document
Text or abstract
Descriptors
Electrolite
Acid
Lactic acid
Muscle
Physiology
Metabolism
Sodium
Etc...
Heading of prefiling

Fig. 33. Microfiches (specimens)
Top: Standard microfiche 105 x 145 mm
Botton: Filmorex card

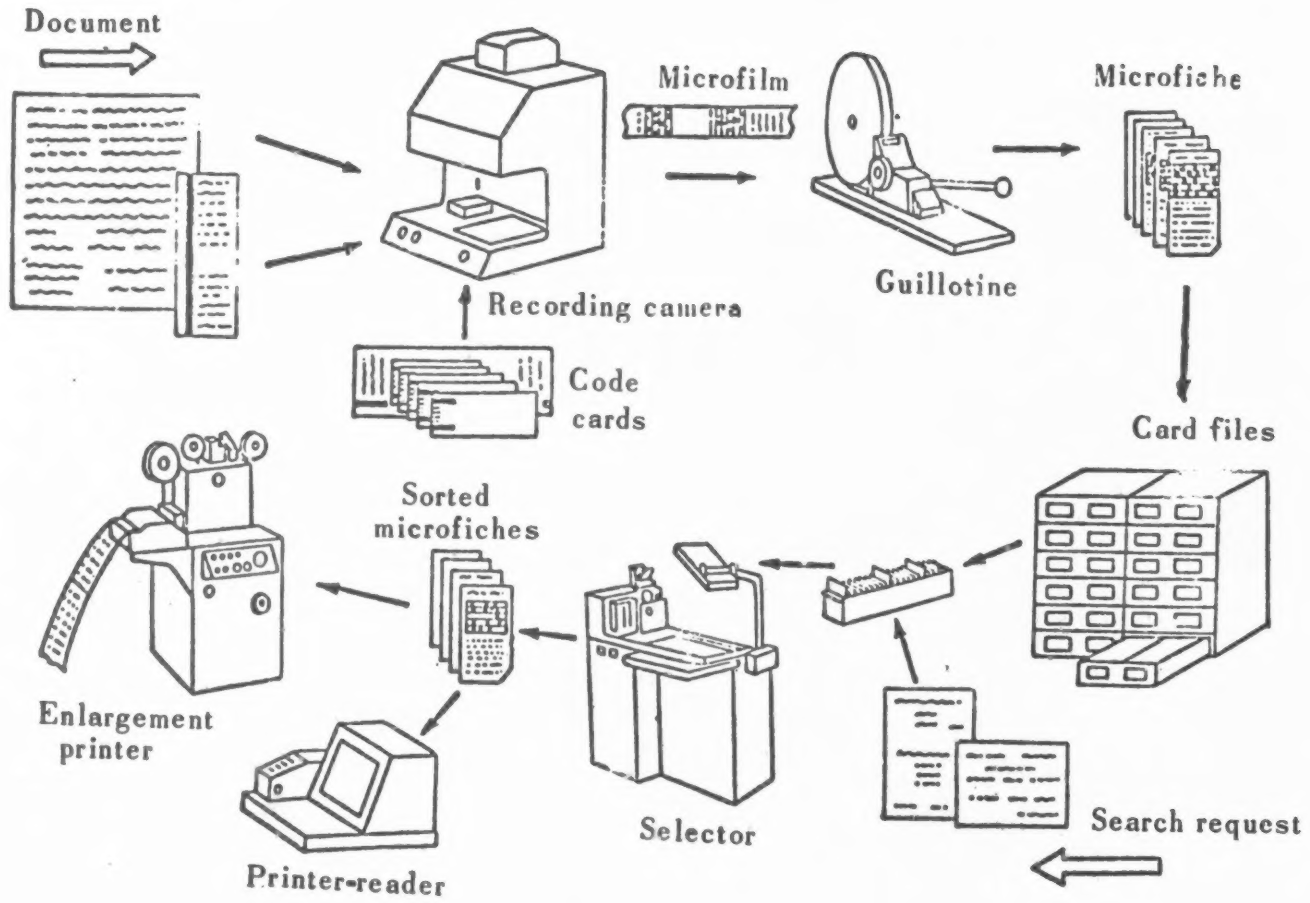

Document
Recording camera
Microfilm
Guillotine
Microfiche
Code cards
Card files
Sorted microfiches
Enlargement printer
Printer-reader
Selector
Search request

Fig. 34. Schematic operation and main units of Filmorex system.

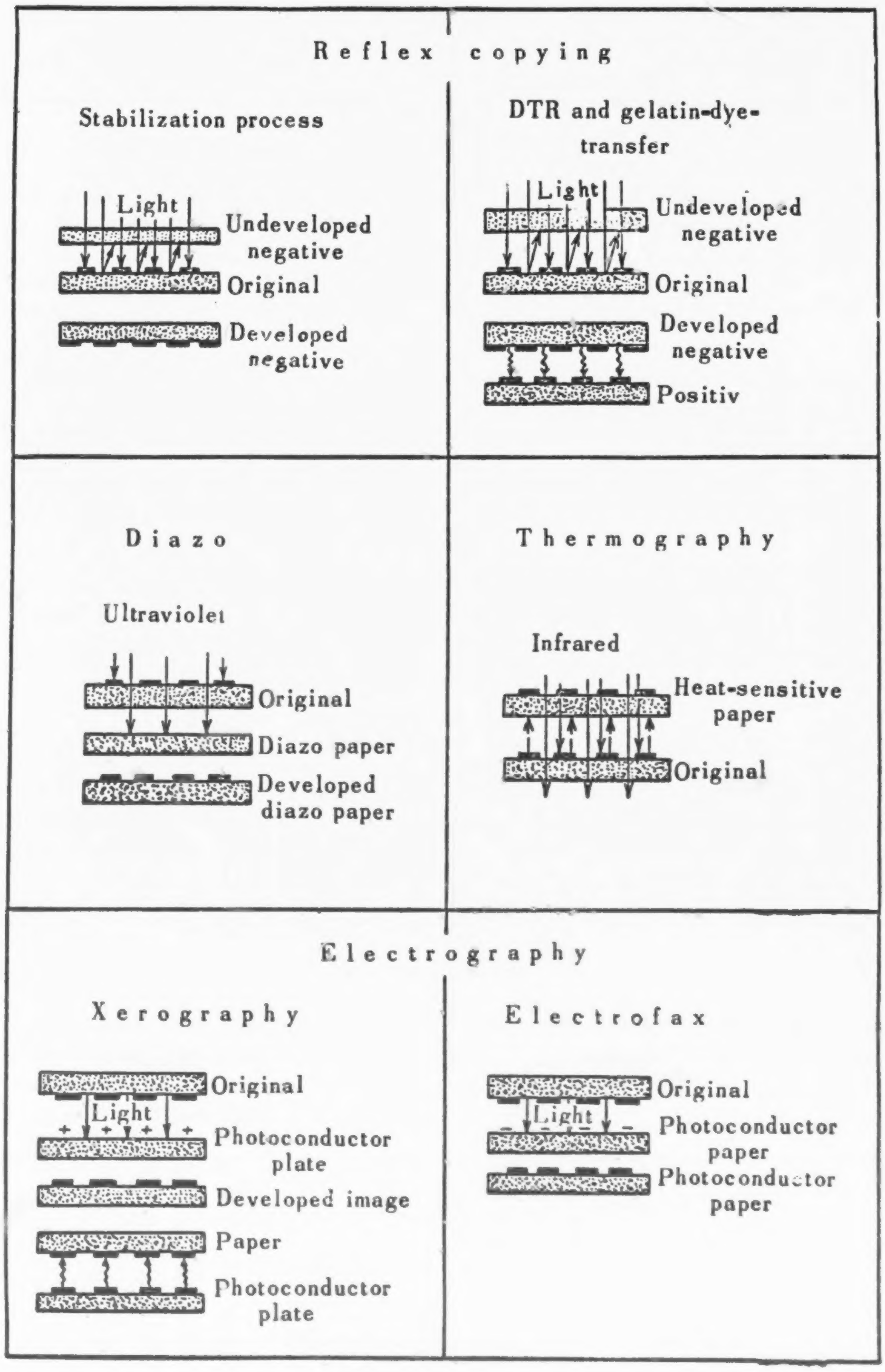

Reflex copying
Stabilization process
Light
Undeveloped negative
Original
Developed negative
DTR and gelatin-dye-transfer
Light
Undeveloped negative
Original
Developed negative
Positiv
Diazo
Ultraviolet
Original
Diazo paper
Developed diazo paper
Thermography
Infrared
Heat-sensitive paper
Original
Electrography
Xerography
Original
Light
Photoconductor plate
Developed image
Paper
Photoconductor plate
Electrofax
Original
Light
Photoconductor paper
Photoconductor paper

Fig. 35. Schematics of the basic copying processes

Advantages	Disadvantages
Reflex copying	
High-contrast copies from line and half-tone (excluding Verifax) originals of any colour (except yellow)	Copies require drying up and tend to darken or fade out with time. Their cost is rather high.
Diazo	
Lasting and most cheap copies from originals of any colour (except blue and violet).	Requires translucent original with image on one side and involves the use of highly odorous chemicals.
Thermography	
The fastest and easiest method of copying of letterpress or typescript images and line drawings. Dry process in lighted room.	Does not copy inscriptions in writing ink and other organic dyes. Copies will darken and become brittle in time.
Xerography	
Lasting copies of excellent quality from textual and line originals of any colour on standard stationery and many other materials. Dry processes in lighted room.	Poor copies from half-tone illustrations and originals with large solid areas. Expensive equipment requiring skilled personnel to operate.
Electrofax	
The basic merits of xerography, plus quality copies from illustrations and originals with large solid areas.	Requires special stationery, from which the image may be easily scratched with metal objects. Some apparatus involve wet processes.

Fig. 36. Relative advantages and disadvantages of the different copying processes.

A

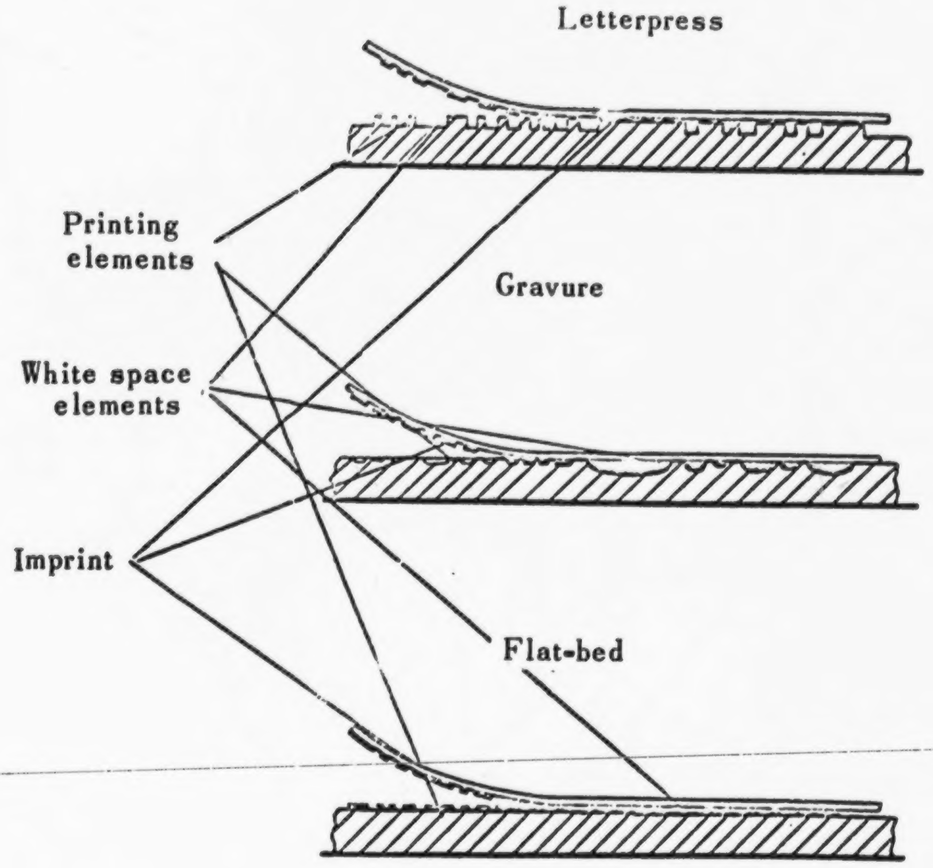

Letterpress
Printing elements
White space elements
Gravure
Imprint
Flat-bed

Fig. 37. Schematics of the different printing methods

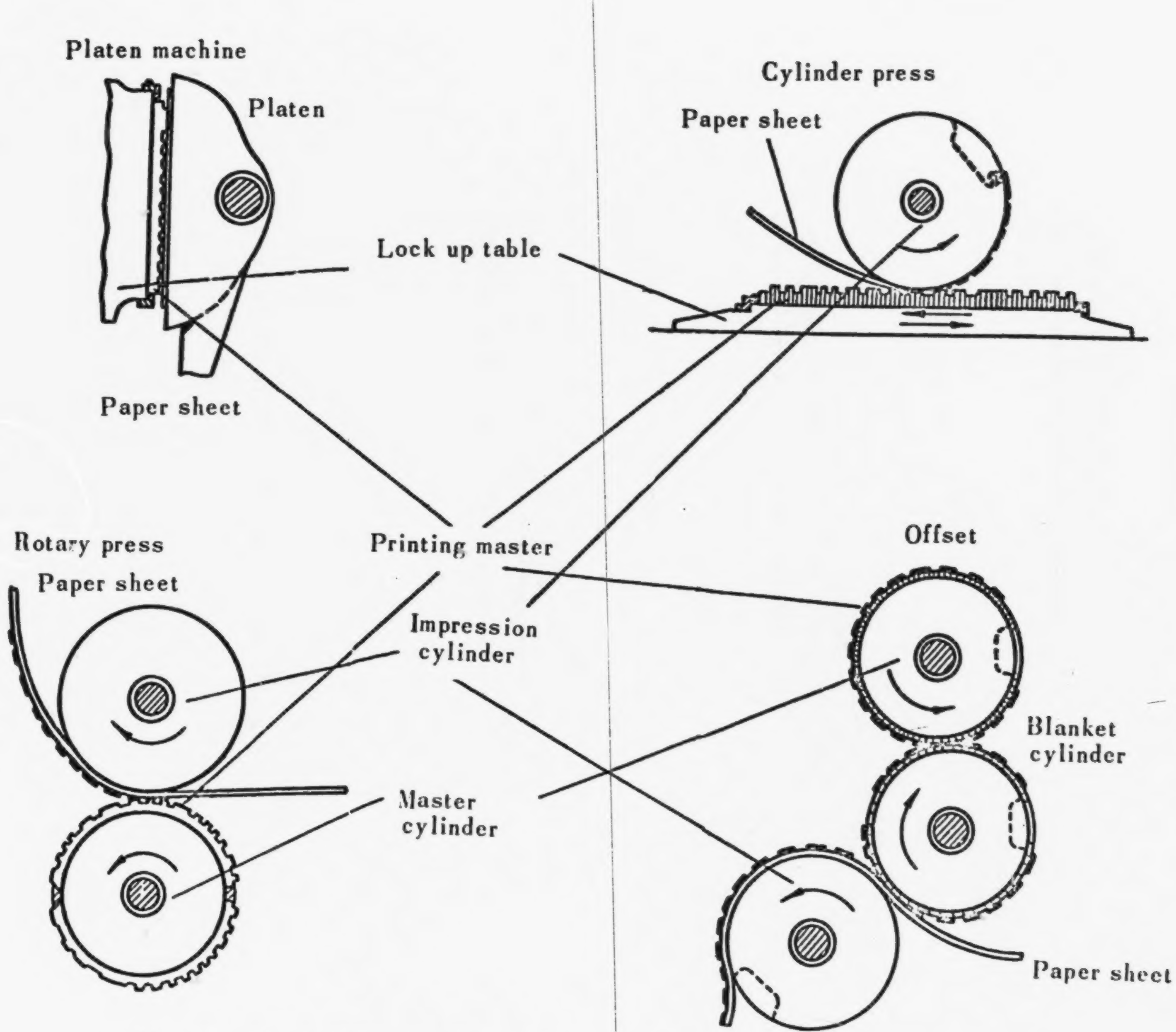
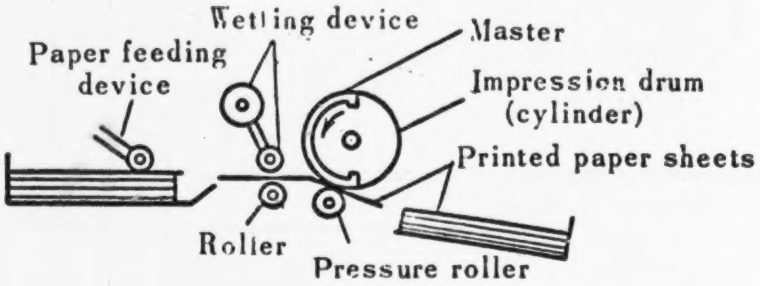
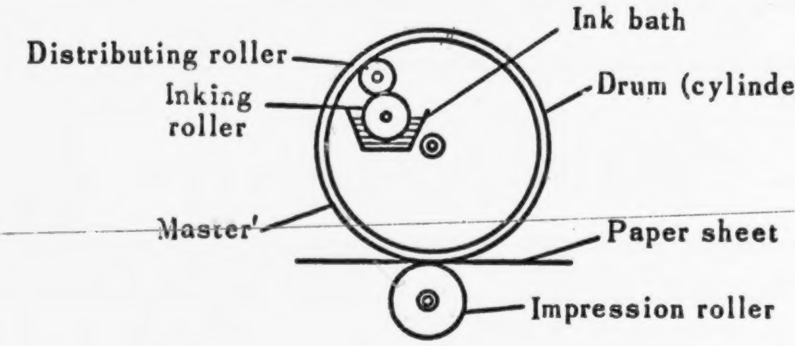

Platen machine
Platen
Paper sheet
Lock up table
Cylinder press
Paper sheet
Printing master
Impression cylinder
Master cylinder
Rotary press
Paper sheet
Offset
Blanket cylinder
Paper sheet

Fig. 38. Block-diagrams of the different printing machine types

Hectograph


Wetting device
Master
Paper feeding device
Impression drum (cylinder)
Printed paper sheets
Roller
Pressure roller

Rotary stencil duplicator


Ink bath
Distributing roller
Drum (cylinder)
Inking roller
Master
Paper sheet
Impression roller

Rotaprint

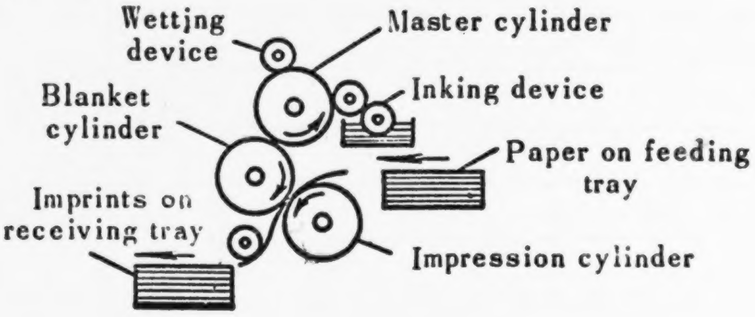

Wetting device
Master cylinder
Blanket cylinder
Inking device
Paper on feeding tray
Imprints on receiving tray
Impression cylinder

Fig. 39. Schematics of operation of the hectograph, rotary stencil duplicator and rotaprint.

Adopted from the book "Razmnozheniye tekhnicheskoy dokumentatsii" (Duplication of technical documents), by V.D. Zasov and V.N. Yurin. Moscow, "Mashinostroyeniye, 1968, p. 129, 135, 137.

Criterion for selection			Typewriter	Automatic typewriter	Hectograph (spirit duplicating)	Hectograph with changeable stamping mechanism	Rotary stencil duplicator with typescript "wax" stencil	Rotary stencil duplicator with electronically cut stencil	Rotaprint (offset duplicating)	Addressing machine
Imprint	Acceptable quality				▩	▩				
	Good quality		▩				▩			▩
	Excelent quality			▩				▩	▩	
Text	Typewritten		▩	▩		▩	▩			▩
	Any kind				▩	▨		▩	▩	
Illustrations	Line				▩		▨	▩	▩	
	Fine line (engineering drawings)							▩	▩	
	Half-tone							▩	▩	
Colour	One colour		▩	▩	▩	▩	▩	▩	▩	▩
	Multicolour	in one run			▩	▨				
		in several runs					▨	▩	▩	
Run	Up to 6 copies		▩	▩						▩
	From 6 to 12		▨	▩	▨	▨				▩
	From 12 to 100			▩	▩	▩				▩
	From 100 to 300			▨	▩	▩	▨	▨	▨	▩
	From 300 to 3000						▩	▩	▩	▩
Stationery	Writing		▨	▨	▩	▩			▨	▩
	Typographic		▩	▩			▨			▩
	Offset						▨	▨	▩	▩
	Note						▩	▩		▩

Fig. 40. Choosing means of document duplication.

Symbols: ▩ recommendable
▨ acceptable

Adopted from: Schöppenthau H. Rationelles Vervilfaltigen. – "Rechentechnik", 1966, Jg. 3, № 6, S. 27.